

EVIDENCE-BASED POLICYMAKING: DESAFIOS E OPORTUNIDADES PARA PORTUGAL

José Eduardo Feio

RESUMO

Nas últimas décadas houve vários avanços metodológicos nas ciências sociais, que são cada vez mais capazes de analisar a existência de relações causais entre uma certa política pública, e.g. rendimento básico universal, e variáveis de interesse, e.g. pobreza. Isto levou ao crescimento de um movimento apelidado de *evidence-based policymaking*, i.e. políticas públicas baseadas na evidência.

Este ensaio começa por analisar métodos das ciências sociais, como, por exemplo, estudos randomizados controlados, que são capazes de fazer uma análise rigorosa da relação entre uma política pública e uma variável de interesse. Por outras palavras, que são capazes de criar evidência sobre se uma política pública está, ou não, a ter o efeito desejado.

De seguida, algumas críticas ao movimento *evidence-based policymaking* são apresentadas. Nomeadamente, limitações à aplicação de evidência em contextos diferentes daqueles em que esta foi recolhida, a importância da definição subjetiva das variáveis de interesse e a necessidade de estudar o processo de implementação de certas políticas públicas.

Este ensaio termina com três sugestões de como Portugal poderia tirar proveito deste movimento. Primeiro, é essencial que haja uma mudança de mentalidades vinda do topo e que os políticos comecem a especificar mais concretamente os objetivos das políticas públicas que escolhem. Segundo, é essencial a criação de instituições, quer públicas quer privadas, que recolham a evidência científica existente e que estudem como esta pode ser aplicada à realidade portuguesa. Terceiro, é importante que se realizem estudos pilotos que permitam, entre outros aspetos, estudar o processo de implementação de políticas públicas.

JOSÉ EDUARDO FEIO
é estagiário no Gabinete da
Deputada ao Parlamento Europeu
Maria Manuel Leitão Marques. É
licenciado com First Class Honours
pela University College London
com um major em Societies e
detentor de um mestrado em
Comparative Social Policy pela
Universidade de Oxford. Durante
os estudos recebeu uma bolsa de
investigação da Fundação Laidlaw,
tendo realizado projetos de
investigação no Instituto Europeu
da University College London
e na LSE IDEAS – o think tank
de política externa da London
School of Economics. Foi também
estagiário Bluebook na Comissão
Europeia. É Membro Honorário
da Associação Portuguesa do
Parlamento Europeu dos Jovens,
associação da qual foi membro
ativo e Presidente entre 2017 e
2018. É atualmente Membro da
Direção da Associação Portuguesa
de Estudos Europeus.

ABSTRACT

In the last decades there have been various methodological advances in the social sciences, which have enhanced researchers' ability to analyze the existence of causal relations between a certain public policy, e.g. universal basic income, and variables of interest, e.g. poverty. This has led to a growing movement under the name of evidence-based policymaking. This essay starts by analyzing methods from the social sciences, such as randomized controlled trials, that can rigorously analyze the relation between a public policy and variables of interest. In other words, they can create evidence about whether a public policy is or not having the desired effect. Then, this essay presents some of the critiques to the evidence-based policymaking movement. Namely, limitations on the application of evidence in contexts different than those in which it was created, the importance of the subjective definition of variables of interest and the need to study the implementation process of public policies. This essay ends with three suggestions of how Portugal could benefit from this movement. First, it is essential that there is a top-down change in mentality and that politicians start being more concrete about the objectives of the public policies they implement. Second, it is essential that we create institutions, either private or public, that collect the available scientific evidence and study how it can be applied to the Portuguese context. Third, it is important that pilot projects take place, these would allow, between other aspects, for the study of the implementation process of public policies.

1.

POLÍTICAS PÚBLICAS BASEADAS NA EVIDÊNCIA

A ideia de que as políticas públicas, i.e. medidas que um governo implementa para resolver problemas, devem ser baseadas em evidência, i.e. factos sobre o efeito de uma política pública numa variáveis de interesse, não é nova e pode parecer até óbvia.

No entanto, a verdade é que durante a grande maior parte do século XX, a evidência sobre o efeito de uma política pública numa certa variável de interesse, como, por exemplo, o efeito da dimensão de turmas escolares nas notas médias dos alunos, não era necessariamente um fator tido em conta no processo político (Bochel and Duncan, 2007). Isto advinha, em grande medida, do facto desta evidência em muitos casos não existir, uma vez que não havia capacidade para recolher dados que permitissem que esta fosse criada. Isto levava a que as políticas públicas fossem guiadas por outros fatores como, por exemplo, argumentos filosóficos, princípios éticos, intuições, interesses de certos grupos e considerações gerais sobre o que era ou não correto fazer (Baldwin et al., 2011).

Este *status quo* é ainda vigente em vários países, mas desde os anos 1990 que se criou um movimento, quer na política quer na academia, que defende a importância da evidência na definição das políticas públicas. Este movimento é normalmente chamado de *evidence-based policymaking*¹, i.e. políticas públicas baseadas na evidência (Parkhurst, 2017).

Este movimento deve muito ao governo do New Labour de Tony Blair (1997 – 2007), no Reino Unido. Na sua campanha e também no seu governo, Tony Blair defendeu que iria “fazer o que funcionava” (Davies et al., 1999). Isto é, iria escolher as políticas públicas baseando-se em evidência científica sobre o seu efeito. Este movimento levou, no Reino Unido, ao crescimento de um ramo das políticas públicas focado em investigar relações causais entre uma certa política pública, e.g. mudanças na forma como as pessoas recebem subsídios, e um certo objetivo de interesse, e.g. aumento da empregabilidade de quem recebe subsídios.

Este ensaio pretende analisar o conceito de *evidence-based policymaking* e a sua aplicabilidade à realidade portuguesa. Este ensaio começa por analisar a capacidade e as limitações de vários métodos das ciências sociais de analisar a relação entre uma política pública e variáveis de interesse. A seguir, este ensaio discute críticas ao movimento *evidence-based policymaking*. Este ensaio termina com uma análise da realidade portuguesa e sugestões sobre passos a tomar.

2.

ANÁLISE DE RELAÇÕES CAUSA-EFEITO ENTRE POLÍTICAS PÚBLICAS E VARIÁVEIS DE INTERESSE

Vários foram os desenvolvimentos metodológicos nas últimas décadas que permitiram que fosse possível a análise causal do efeito de uma política pública em variáveis de interesse. Podemos dividir estes métodos em: estudos controlados randomizados (*randomised control trials, RCT*); estudos não randomizados (Greve, 2017).

Nos estudos controlados randomizados os participantes são distribuídos aleatoriamente entre dois grupos, um grupo que recebe a intervenção e outro que serve de controlo. Nos métodos não randomizados, pode existir na mesma um grupo de controlo, mas a distribuição entre grupo de intervenção e grupo de controlo não é aleatória.

1 A literatura deste tópico encontra-se escrita maioritariamente em inglês. Por isso, quando há tradução de termos como *randomised control trials* e *evidence-based policymaking*, é sempre indicado o termo em inglês.

2.1.

ESTUDOS CONTROLADOS RANDOMIZADOS (RCT)

Os estudos controlados randomizados, em inglês *randomised control trials*, são considerados o método mais rigoroso para analisar se existe uma relação de causa-efeito entre uma certa política pública e uma variável de interesse (Pearce and Raman, 2014). Este método foi desenvolvido nas ciências da saúde para a análise do efeito de medicamentos e tem sido cada vez mais aplicado nas ciências sociais.

Este método requer que se recrute participantes a partir da população de interesse, e.g. população portuguesa, e que estes sejam distribuídos aleatoriamente entre dois grupos. Um dos grupos, o grupo de controlo, não recebe qualquer intervenção. O outro grupo, o grupo da intervenção, estará sob o efeito da medida que se pretende estudar (Haynes et al., 2012).

A grande vantagem deste método é que, como os participantes são distribuídos aleatoriamente entre dois grupos, pode-se assumir que quaisquer fatores, para além da intervenção que se está a estudar, que possam afetar o resultado, estão distribuídos de forma igual pelos dois grupos. Ou seja, qualquer diferença no valor da variável de interesse entre o grupo de controlo e o grupo de intervenção vai ser devido à política pública que se está a estudar (Cartwright, 2007).

Este método de análise de políticas públicas, i.e. de avaliar se uma política pública tem efeito numa variável de interesse, recebeu bastante reconhecimento de instituições oficiais. Documentos do Governo Britânico, por exemplo, falam deste método com “a melhor forma de saber se uma política pública está a funcionar” (traduzido a partir de Haynes et al. (2012)).

Greve (2017) apresenta um caso de estudo que ilustra bem o valor deste método. Na década de 1970 em New Jersey, nos Estados Unidos, foi iniciado um programa, i.e. uma política pública, que pretendia reduzir o número de crimes cometidos por jovens que viviam em condições desfavoráveis, i.e. variável de interesse. Neste programa, jovens em risco iam a prisões conhecer presos que lhes falavam da sua vida e de como chegaram à prisão. Alguns estudos iniciais que não usaram métodos randomizados indicaram níveis de sucesso na ordem dos 90%. Isto levou a que este tipo de programas se espalhasse pelos Estados Unidos. No entanto, mais recentemente, estudos controlados randomizados indicaram que estes programas não tinham nenhum efeito na variável de interesse, i.e. redução da incidência de crime (Petrosino et al., 2003).

Por outras palavras, os resultados iniciais, que levaram à expansão destes programas por todos os Estados Unidos, estavam afetados por outros fatores que

não o programa. Por exemplo, é possível que os alunos que aceitavam ir a estas apresentações eram quem já à partida tinha menos probabilidade de cometer crimes. Quando se randomizou os participantes, ou por outras palavras, quando se controlou para todos os fatores que podiam afetar a variável de interesse para além da intervenção, verificou-se que não havia uma diferença significativa entre os jovens que estavam presentes nestas palestras e os que não estavam.

Este exemplo é bastante ilustrativo dos benefícios dos estudos controlados randomizados. Caso um estudo controlado randomizado tivesse sido feito mais cedo, teria sido possível observar que não havia uma relação causa-efeito entre estes programas e a redução da criminalidade jovem. Isto teria evitado que se investisse tempo, dinheiro e recursos humanos numa medida que não tem o efeito desejado.

Mas se os estudos controlados randomizados são assim tão bons a definir se há ou não uma relação de causa-efeito entre uma política pública e uma variável de interesse, i.e. a recolher evidência sobre políticas públicas, porque não são mais usados?

Primeiro, os estudos controlados randomizados requerem um grande investimento de recursos humanos e financeiros (Hawkins, 2016). Segundo, estes são muito demorados, o que não se adequa ao tempo que os políticos têm para tomar decisões, que tende a ser bastante curto (O'Reilly and Vingilis, 2018). Terceiro, há vários problemas éticos relacionados com os estudos controlados randomizados, pois estes implicam que não se dê uma intervenção, que se acha necessária para resolver um problema, a um grupo de pessoas. Quarto, há intervenções que não podem ser estudadas com estudos controlados randomizados, e.g. mudanças em políticas de impostos, pois estas só funcionam quando são aplicados a toda a população e por isso não é possível criar um grupo de controlo e um grupo de intervenção (Hanley et al., 2016).

2.2.

ESTUDOS NÃO RANDOMIZADOS

Os limites dos estudos controlados randomizados, especialmente questões de recursos e de tempo, levaram ao desenvolvimento de outros métodos de análise, baseados em análises estatísticas avançadas. Diferentes métodos foram desenvolvidos para tirar partido de diferentes oportunidades e especificidades das políticas públicas que se pretende analisar. Dou aqui três exemplos: experiências naturais, designs em regressão descontínua e séries temporais interrompidas.

Experiências naturais são experiências que tentam tirar partido de um evento exógeno que causa uma distribuição que se pode assumir como quase aleatória

da população entre dois grupos. Por exemplo, quando o Furacão Katrina afetou Nova Orleães levou a que 50% dos presos não pudessem voltar ao seu bairro de residência quando fossem libertados. Kirk (2009) aproveitou esta situação para explorar se o facto de os presos voltarem para o seu bairro de residência afetava a sua probabilidade de reincidirem, tendo verificado uma redução de 15% de reincidência para presos realojados fora do seu bairro de residência.

Temos também estudos com um design em regressão descontínua, *regression discontinuity designs*. Estes são usados quando há um limiar, por exemplo a idade, que define se um sujeito está ou não sob o efeito de uma medida. A ideia por de trás deste método é que os sujeitos perto desse limiar, tanto do lado da intervenção como do lado que não está sob a intervenção, são comparáveis. MacDonald et al. (2016) usou este método para estudar o efeito da presença de equipas de polícia privadas na incidência de crime. Para isso, analisou bairros dentro da área do campus da Universidade da Pensilvânia (que tem patrulhas privadas) e bairros perto deste campus, mas que não tinham este tipo de patrulhas, verificando uma redução da incidência de crime em áreas com patrulhas de polícia privadas.

Um outro método são estudos de séries temporais interrompidas, *interrupted time series* (Shadish and Cook, 2009). Este método serve para explorar efeitos relacionados com mudanças repentinas, que podem ser devido à implementação de uma lei, a uma mudança em políticas públicas ou até desastres naturais. Este método baseia-se na comparação entre a situação pré e pós mudança e requer o uso de várias variáveis de controlo e ajustes estatísticos avançados. Humphreys and McLellan (2011) analisaram a introdução de uma lei na Flórida que passava a dar imunidade a quem usasse força letal contra outra pessoa, caso isto ocorresse em defesa própria. Usando uma base de dados alargada com muitas variáveis a servir de controlo, eles analisaram o número de homicídios antes e depois desta lei e verificaram um aumento de cerca de 24.7% no número mensal de homicídios.

Estes são apenas três exemplos, mas muitos outros métodos foram desenvolvidos, como por exemplo *propensity score matching* e *difference in difference designs* (Cook et al., 2002). Cada um destes métodos é desenvolvido para aproveitar certas circunstâncias específicas em que uma política pública é aplicada. A grande diferença entre os estudos controlados randomizados e estes métodos é a introdução de uma maior propensão para o erro. Isto acontece porque as condições experimentais não são tão rigorosas como nos estudos controlados randomizados. Em específico, não há uma distribuição controlada pelos investigadores dos participantes aleatoriamente entre um grupo de intervenção e um grupo de controlo. Por isso, há sempre o risco de haver fatores que não são controlados pela experiência e que afetam a variável de interesse, para além da política pública que se está a estudar.

Mas várias análises destes métodos comprovam que os resultados de estudos não randomizados podem ser tão rigorosos como os de estudos controlados randomizados (Cook et al., 2008, Shadish and Cook, 2009). Isto acontece quando a escolha do método de análise é adequada à situação e quando são utilizadas variáveis de controlo e métodos estatísticos avançados de acordo com a variável de interesse. Quando isto acontece, os resultados dos estudos randomizados e dos estudos não randomizados são semelhantes. Os estudos não randomizados têm então a vantagem de poderem ser tão rigorosos como os randomizados, mas requerendo menos tempo, recursos humanos e financeiros.

3.

PARA ALÉM DA CAUSA-EFEITO: CRÍTICAS A *EVIDENCE-BASE POLICYMAKING*

As seções anteriores analisaram vários métodos que são capazes de avaliar a existência de uma relação causa-efeito entre uma política pública e uma variável de interesse. Por outras palavras, é possível criar evidência sobre a eficácia de políticas públicas.

Mas isto não significa que podemos seguir cegamente a evidência criada a partir destes métodos. Este capítulo analisa algumas das críticas feitas ao movimento *evidence-based policymaking* e cuidados que se deve ter quando se considera a importância de “fazer o que funciona”.

3.1.

VALIDADE INTERNA VS VALIDADE EXTERNA

Uma das grandes limitações dos métodos analisados nas seções anteriores relaciona-se com os conceitos de validade interna e validade externa.

Um método tem validade interna se os resultados que dele advêm forem uma boa aproximação aos resultados verdadeiros que existem na população que se está a estudar. Um método tem validade externa se os resultados que dele advêm poderem ser aplicados a um contexto diferente daquele que está a ser estudado, e.g. um outro país (McDermott, 2011).

Os métodos analisados em cima são muito fortes em termos de validade interna. Isto é, conseguem ter resultados que se aproximam do valor real do efeito de uma certa política pública em variáveis de interesse (Haynes et al., 2012).

Mas a sua validade externa não é forte. Isto é, não é claro à partida que os resultados que advêm do uso destes métodos possam ser aplicados a outras

populações para além da que está a ser estudada (Cartwright, 2007). Isto advém do facto de haver fatores, como por exemplo a cultura, o envolvimento institucional e a estrutura do estado, que variam bastante de país para país.

Mesmo em estudos randomizados não é possível controlar alguns destes fatores. Num estudo controlado randomizado feito no Reino Unido, por exemplo, tanto o grupo de intervenção como o grupo de controlo vão ser afetados pela cultura e estruturas institucionais britânicas. Um estudo controlado randomizado não permite saber qual o efeito destes fatores e a relevância de mudanças nestes, e.g. culturas diferentes, para a relação causa-efeito entre uma política pública e uma variável de interesse. É possível que, por exemplo, por questões culturais, uma certa política pública, como o uso de campanhas de sensibilização sobre os efeitos negativos do consumo de álcool, tenham um efeito positivo numa variável de interesse, e.g. redução do consumo, num país e não noutra (O'Reilly and Vingilis, 2018).

Estes fatores são difíceis de estudar e de quantificar e, por isso, limitam aplicabilidade dos resultados obtidos através destes métodos a outras realidades.

3.2.

DEFINIÇÃO E CONCEPTUALIZAÇÃO DOS PROBLEMAS E OBJETIVOS DAS POLÍTICAS PÚBLICAS

Uma das grandes críticas ao movimento *evidence-based policymaking* é feita por investigadores que trabalham numa perspetiva construtivista, i.e. estão interessados em como os atores sociais constroem e conceptualização a realidade à sua volta (Bacchi and Goodwin, 2016).

O movimento *evidence-based policymaking* tende a assumir que os problemas a resolver e os objetivos das políticas públicas são objetivos. Mas a verdade é que há maneiras muito diferentes de definir certos problemas e, por consequência, de definir as variáveis de interesse (Bacchi, 2015). Um problema como o consumo de drogas pode ser definido como um problema criminal, pelo uso de uma substância ilícita, como nos EUA, ou como um problema de saúde pública, por ser uma ação que causa danos aos indivíduos que a fazem, como em Portugal. E, dependendo de como se define o problema, a variável de interesse vai ser diferente, e.g. consumo de drogas ilícitas per capita vs número de infetados com HIV/SIDA.

Por causa disto, dizer que uma política pública “funciona” pode ser falacioso, porque a definição do que funciona, i.e. da variável de interesse que é afetada pela política pública, pode ser uma com a qual não concordemos.

3.3.

COMPLEXIDADES NA IMPLEMENTAÇÃO
DE POLÍTICAS PÚBLICAS

Uma outra crítica ao movimento *evidence-based policymaking* tem a ver com o facto de este ignorar a complexidade da implementação de certas políticas públicas.

O foco na análise da relação causa-efeito entre uma política pública e uma variável de interesse ignora a importância do processo de implementação para o sucesso de políticas públicas (Greve, 2017). Algumas políticas públicas têm uma implementação simples, como por exemplo mudanças nos impostos. Outras políticas públicas, como, por exemplo, mudanças na estrutura do sistema nacional de saúde, são muito complexas e requerem que se coordene vários atores. Esta complexidade de implementação leva a que muitos governos prefiram fazer projetos pilotos direcionados a estudar a melhor forma de implementar uma medida, em vez de se focarem numa análise de causa-efeito (Bochel and Duncan, 2007).

Isto não implica que a análise da relação causa-efeito entre uma política pública e uma variável de interesse deva ser negligenciada, mas demonstra que muitas vezes esta não providencia toda a informação necessária para a implementação de políticas públicas. Estas análises de implementação podem complementar análises causais, algo que é especialmente importante quando se está a importar uma medida de um contexto diferente (Bochel and Duncan, 2007).

4.

OPORTUNIDADES E SUGESTÕES PARA PORTUGAL

Será bastante consensual dizer que o movimento *evidence-based policymaking* ainda não chegou a Portugal. Isto é, basear as políticas públicas em evidência sobre o seu efeito em variáveis de interesse ainda não é uma prática comum. Isto revela-se no discurso político, na forma como as políticas públicas são apresentadas e como elas são discutidas no espaço público, mas também na falta de instituições públicas e privadas que se dedicam à recolha de evidência científica que possa ser utilizada para informar políticas públicas.

Isto pode ser visto como uma oportunidade de observar o que se passou em outros países, como no Reino Unido, e através disto introduzirmos os aspetos benéficos deste movimento, estando sempre conscientes dos seus limites.

Há três opções que, considerando o discutido nos capítulos anteriores, podem ser sugeridas para Portugal. De um modo geral, precisamos de uma mudança

de mentalidade que tem que vir do topo, como aconteceu com Tony Blair. Isto implica não só uma mudança mais abstrata em termos do discurso que os governantes têm em torno das políticas públicas, mas também medidas concretas como a especificação de objetivos claros para todas as políticas públicas que são apresentadas. O primeiro passo na análise de políticas públicas é a especificação dos objetivos das mesmas, pois isto permite definir quais são as variáveis de interesse. Sem isto, é impossível analisar se as políticas tiveram ou vão ter o efeito desejado. E mais, a especificação dos objetivos traz valor acrescentado ao processo político por si mesma, pois aumenta a transparência do mesmo.

Segundo, é necessário criar instituições que recolham evidência científica e que analisem como esta pode ser aplicada ao contexto português. No Reino Unido existem vários centros de investigação chamados “What Works Centers” (Bristow et al., 2015). Estes centros de investigação são financiados por recursos públicos e privados e têm como objetivo recolher, criar e partilhar evidência científica que possa informar políticas públicas. Estes centros produzem, entre outras coisas, sínteses da evidência científica em várias áreas, analisam políticas públicas e apoiam investigadores e decisores políticos nas suas ações. A criação de centros deste género em Portugal seria uma maneira muito eficaz de lançar o movimento *evidence-based policymaking* no nosso país.

Terceiro, seria importante que, sempre que possível, se realize projetos piloto das políticas públicas que se pretende implementar. Estes serviriam para estudar o processo de implementação das mesmas, mas poderiam também ser usados para fazer análises preliminares dos efeitos das políticas em variáveis de interesse e para preparar a recolha de dados para estudos mais alargados.

Com estas três opções, que podem ser implementadas em conjunto ou em separado, seria possível que Portugal começasse a aderir a este movimento. Desde que fosse feito com consciência das limitações do mesmo, isto poderia trazer vários benefícios ao nosso país, como uma melhor alocação dos recursos através de melhorias no processo de escolha das políticas públicas a implementar.

5. CONCLUSÃO

Em conclusão, o movimento *evidence-based policymaking* pode trazer muitos benefícios para Portugal. O uso de métodos de investigação para analisar se existe uma relação causa-efeito entre uma política pública e variáveis de interesse pode permitir o aumento da eficiência das ações do estado e a poupança de recursos, quer humanos quer financeiros.

Para isto, é essencial que estejamos conscientes dos limites deste movimento. Nomeadamente, limitações à aplicação de evidência em contextos diferentes daqueles em que esta foi recolhida, a importância da definição subjetiva das variáveis de interesse e a necessidade de estudar o processo de implementação de certas políticas públicas.

Este ensaio fez três sugestões de como Portugal poderia tirar proveito deste movimento. Primeiro, é essencial que haja uma mudança de mentalidades vinda do topo e que os políticos comecem a especificar mais concretamente os objetivos das políticas públicas que escolhem. Segundo, é essencial que se crie instituições, quer públicas quer privadas, que recolham a evidência científica existente e que estudem como esta pode ser aplicada à realidade portuguesa. Terceiro, é importante que se realizem cada vez mais estudos piloto, que permitam, entre outros aspetos, estudar o processo de implementação de políticas públicas.

BIBLIOGRAFIA

- BACCHI, C. 2015. Problematizations in Alcohol Policy: WHO's "Alcohol Problems". *Contemporary Drug Problems*, 42, 130-147.
- BACCHI, C. & GOODWIN, S. 2016. *Poststructural policy analysis: A guide to practice*, Springer.
- BALDWIN, R., BALDWIN, R., CAVE, M., CAVE, M., LODGE, M. & LODGE, M. 2011. *Understanding Regulation: Theory, Strategy, and Practice*, Oxford University Press.
- BOCHEL, H. & DUNCAN, S. 2007. *Making policy in theory and practice*, Policy Press.
- BRISTOW, D., CARTER, L. & MARTIN, S. 2015. Using evidence to improve policy and practice: the UK What Works Centres. *Contemporary social science*, 10, 126-137.
- CARTWRIGHT, N. 2007. Are RCTs the Gold Standard? *BioSocieties*, 2, 11-20.
- COOK, T. D., CAMPBELL, D. T. & SHADISH, W. 2002. *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin Boston, MA.
- COOK, T. D., SHADISH, W. R. & WONG, V. C. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within study comparisons. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 27, 724-750.
- DAVIES, H. T., NUTLEY, S. M. & SMITH, P. C. 1999. Editorial: What works? The role of evidence in public sector policy and practice.
- GREVE, B. 2017. *Handbook of social policy evaluation*, Northampton, MA, Edward Elgar Pub.
- HANLEY, P., CHAMBERS, B. & HASLAM, J. 2016. Reassessing RCTs as the 'gold standard': synergy not separatism in evaluation designs. *International Journal of Research & Method in Education*, 39, 287-298.
- HAWKINS, A. J. 2016. Realist evaluation and randomised controlled trials for testing program theory in complex social systems. *Evaluation*, 22, 270-285.
- HAYNES, L., GOLDACRE, B. & TORGERSON, D. 2012. Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials| Cabinet Office.
- HUMPHREYS, K. & MCLELLAN, A. T. 2011. A policy oriented review of strategies for improving the outcomes of services for substance use disorder patients*. Oxford, UK.
- KIRK, D. S. 2009. A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina. *American Sociological Review*, 74, 484-505.
- MACDONALD, J. M., KLINK, J. & GRUNWALD, B. 2016. The effect of private police on crime: evidence from a geographic regression discontinuity design. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 831-846.
- MCDERMOTT, R. 2011. Internal and external validity. *Cambridge handbook of experimental political science*, 27-40.
- O'REILLY, R. & VINGILIS, E. 2018. Are Randomized Control Trials the Best Method to Assess the Effectiveness of Community Treatment Orders? *Administration and Policy in Mental Health and Mental Health Services Research*, 45, 565-574.

-
- PARKHURST, J. O. 2017. *The politics of evidence : from evidence-based policy to the good governance of evidence*, Abingdon, Oxon, Routledge, Taylor & Francis Group.
- PEARCE, W. & RAMAN, S. 2014. The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy sciences*, 47, 387-402.
- PETROSINO, A., TURPIN-PETROSINO, C. & BUEHLER, J. 2003. Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review of the Randomized Experimental Evidence. *The ANNALS of the American Academy of Political and Social Science*, 589, 41-62.
- SHADISH, W. R. & COOK, T. D. 2009. The Renaissance of Field Experimentation in Evaluating Interventions. *Annu. Rev. Psychol.*, 60, 607-629.