

Título do capítulo	CAPÍTULO 4 – CONECTANDO PESQUISA A GESTÃO MUNICIPAL: AVALIAÇÕES DE IMPACTO INFLUENCIAM A FORMAÇÃO DE POLÍTICA PÚBLICA?
Autores(as)	Diana Moreira Juan Francisco Santini
DOI	http://dx.doi.org/10.38116/978-65-5635-032-5/capitulo4

Título do livro	POLÍTICAS PÚBLICAS E USOS DE EVIDÊNCIAS NO BRASIL: CONCEITOS, MÉTODOS, CONTEXTOS E PRÁTICAS
Organizadores(as)	Natália Massaco Koga Pedro Lucas de Moura Palotti Janine Mello Maurício Mota Saboya Pinheiro
Volume	-
Série	-
Cidade	Brasília
Editora	Instituto de Pesquisa Econômica Aplicada (Ipea)
Ano	2022
Edição	1ª
ISBN	978-65-5635-032-5
DOI	http://dx.doi.org/10.38116/978-65-5635-032-5

© Instituto de Pesquisa Econômica Aplicada – Ipea 2022

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos). Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

CONECTANDO PESQUISA A GESTÃO MUNICIPAL: AVALIAÇÕES DE IMPACTO INFLUENCIAM A FORMULAÇÃO DE POLÍTICA PÚBLICA?¹

Diana Moreira²
Juan Francisco Santini³

1 INTRODUÇÃO

O processo pelo qual políticas são escolhidas e postas em prática é complexo. Envolve vários atores com interesses diversos, muitas vezes com objetivos que pouco se relacionam com o bem-estar do público atingido pela política. Mas mesmo quando a política pública é concebida com a melhor das intenções, objetivando sanar um problema do público, a solução não é facilmente atingida.⁴ Apenas a boa intenção não faz uma boa medicina. A boa medicina requer uma série de passos que, em conjunto, levam ao progresso e à melhora do paciente.

Antes de propor o tratamento, consulta-se o que já comprovadamente funcionou em uma parcela grande de pacientes e se tem conhecimento dos eventuais efeitos colaterais. Uma vez que o tratamento é dado, o acompanhamento continua, observando-se erros de percurso, alterando-se o tratamento até que o problema seja sanado ou ao menos suavizado. Nem sempre existe um tratamento comprovado, e a médica trata o paciente de forma ainda mais cautelosa com doses menores e maior acompanhamento, às vezes até construindo essa comprovação do tratamento. A resolução de problemas depende desse processo sistemático de busca de acertos e erros de tentativas passadas, e contínuo acompanhamento de acertos e erros futuros. A vida do gestor público é bem mais difícil, mas a formulação de política pública poderia funcionar assim, de forma integrada com evidência empírica.

1. Agradecemos os comentários de Diana Kaplan Barbosa e Richard Martins da Cunha, que contribuíram para melhoria deste capítulo. Em conjunto com Diana Moreira e Juan Francisco Santini, Gautam Rao (Harvard University e National Bureau of Economic Research – NBER) e Jonas Hjort (Columbia University, NBER, The Bureau for Research and Economic Analysis of Development – Bread e Centre for Economic Policy Research – CEPR) são também autores do artigo original em inglês no qual este capítulo se baseia. Qualquer referência a este capítulo deve citar o artigo original (Hjort *et al.*, 2021). Erros nas análises adicionais apresentadas, adaptações e tradução devem ser atribuídos apenas aos autores deste capítulo.

2. Professora assistente na Universidade da Califórnia, Davis, Estados Unidos. *E-mail*: <dsmoreira@ucdavis.edu>.

3. Pesquisador na Innovations for Poverty Action (IPA). *E-mail*: <jsantini@poverty-action.org>.

4. Além dos objetivos dos agentes envolvidos na tomada de decisão, existem, claro, restrições. Desde a falta de recursos financeiros e rigidez de processos na máquina pública até a falta de pessoal qualificado. Este capítulo procura discutir se, na prática, a falta de conhecimento da efetividade das diferentes soluções é de fato uma restrição relevante adicional.

Existem diversas formas pelas quais evidências empíricas podem entrar no processo de decisão, desde o diagnóstico do problema até a elaboração, o refinamento e a implementação de potenciais soluções. O foco deste capítulo consiste no uso, por parte de gestores públicos municipais, de evidência advinda de avaliação de impacto quanto à efetividade de políticas públicas. Note que esse enfoque acaba por gerar diferenças com relação a usos de pesquisa debatidos em outros contextos.

A primeira diferença é que nossos resultados são relevantes para políticas públicas municipais. O acesso, as instituições e o nível de polarização ideológica tendem a variar com nível de governo e, infelizmente, é difícil prever se tais resultados se aplicariam para a esfera federal, que é o foco principal dos demais capítulos deste livro. A segunda diferença diz respeito ao tipo de evidência empírica da qual tratamos: avaliação de impacto. Esta consiste em uma análise empírica testando uma hipótese de causa e efeito, tendo um critério previamente estabelecido para confirmar ou rejeitar a hipótese. Isso exclui, portanto, uma camada grande de pesquisas, como aquelas que selecionam casos de sucesso para serem estudados a fundo, ou análises que observam apenas beneficiários de uma política antes e depois da implementação, ou análises correlacionais entre a existência de uma política e o seu objetivo que não usam métodos para isolar os efeitos de outras causas prevalentes no ambiente onde são estudadas. Este capítulo não oferece aprendizados quanto a esses outros tipos de pesquisa.

Estudos de avaliação de impacto têm um potencial enorme de revolucionar a prática de políticas públicas. Bilhões de reais são gastos anualmente em intervenções concebidas para reduzir a pobreza, aumentar a produtividade, melhorar os aprendizados nas escolas, reduzir a mortalidade infantil, entre outros. Entender se uma intervenção atinge ou não seus objetivos, e porque sim e porque não, é crucial para a prestação de contas e o uso eficiente dos recursos. Estudos que avaliam rigorosamente o impacto de políticas públicas não só permitem descobrir os efeitos de tais políticas, mas também permitem ajustar o percurso caso o caminho inicial destas se mostre inadequado. O aprendizado com respeito à efetividade de políticas tem crescido muito nas últimas décadas. Apenas no registro da American Economic Association, desde 2013 iniciou-se mais de 4,5 mil estudos de avaliação de impacto. Indiscutivelmente, estudos de avaliação de impacto não são a panaceia e estão sujeitos, por exemplo, a diferentes questionamentos de validade externa e viés de publicação. Porém, este tipo de estudo permite revelar êxitos e, principalmente, fracassos, o que é fundamental para formular políticas que realmente progridam.

Conectar pesquisa com a gestão pública no Brasil parece talvez um sonho distante. O gráfico 1A mostra que os prefeitos raramente utilizam as universidades e as instituições de pesquisas como fontes para informar-se sobre uma política

pública.⁵ Entretanto, o gráfico 1B oferece uma visão promissora, pois indica que universidades e instituições de pesquisas foram indicadas pelos prefeitos entrevistados como as fontes mais confiáveis (diferença para outras fontes estatisticamente significantes) para obtenção de informação.

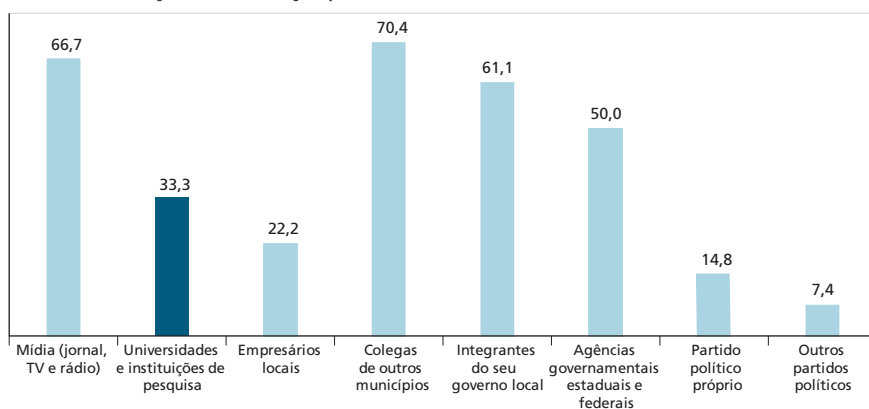
Apesar do potencial das avaliações de impacto de mudar a prática de formular política pública, o quanto de fato muda ou pode mudar depende de fatores adicionais. Até que ponto os líderes políticos estão interessados e abertos a pesquisas de avaliações de impacto? E, na medida em que “consumem” pesquisas, eles conseguem agir com base nas novas descobertas? Considerando o volume enorme de estudos e aprendizados “comprovados”, o entendimento dessas questões é fundamental para que essa capacidade em potencial seja realizada: se os líderes políticos valorizam essas pesquisas; se isso muda suas crenças sobre a efetividade de políticas; e se os líderes implementam políticas públicas a que, de outra forma, não teriam acesso, em resposta às novas descobertas científicas. Em suma, a falta de (acesso a) informações provindas de pesquisas de avaliações de impacto é uma restrição na tomada de decisão de políticas públicas?

Neste capítulo, damos um primeiro passo para responder a essas perguntas, fornecendo evidências de dois experimentos. Em parceria com a Confederação Nacional de Municípios (CNM) no Brasil, a pesquisa descrita a seguir atingiu prefeitos(as) e gestores por todo o Brasil, em um total de 2.150 municípios. Faz-se importante notar que grande parte das análises aqui apresentadas apareceu originalmente no artigo publicado na revista acadêmica *American Economic Review* e vários trechos foram traduzidos e adaptados do artigo original de Hjort *et al.* (2021).

GRÁFICO 1

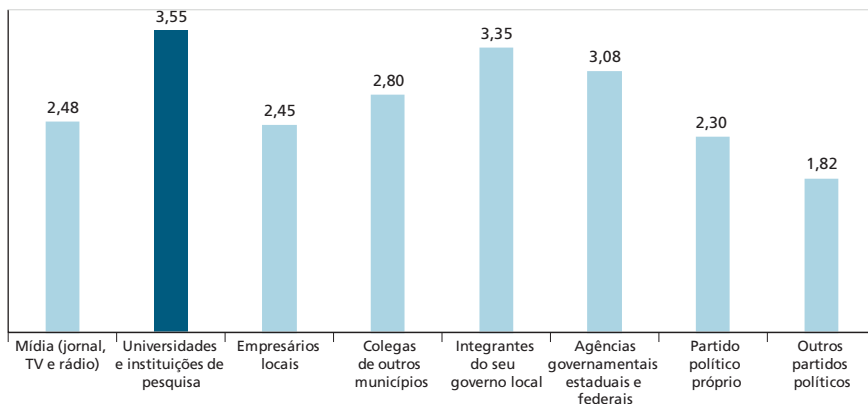
Utilização e grau de confiança em fontes de informação

1A – Obtenção de informação por fonte (%)



5. O gráfico 1 baseia-se em um estudo de caso com setenta prefeitos.

1B – Grau de confiança por fonte



Elaboração dos autores.

Obs.: 1. O gráfico 1A mostra respostas de prefeitos(as) à seguinte pergunta: "De modo geral, a partir de quais fontes você aprende sobre uma nova política ou obtém novas informações sobre uma política já existente (por exemplo: que ela não funciona ou fica sabendo de uma melhor forma de implementá-la)? Por favor, selecione mais de uma opção caso a informação provenha de diferentes fontes". O gráfico 1B mostra respostas de prefeitos(as) à seguinte pergunta: "Para cada um dos seguintes grupos/instituições, por favor, indique o quanto você confia neles como fonte de informação relevante para formulação de políticas locais", sendo: (1) nada; (2) não muito; (3) bastante; e (4) muito.

2. Quando se trata de obtenção de informações, as universidades e instituições de pesquisa ficaram acima somente de empresários locais e partidos políticos, que são as fontes menos confiáveis. A pesquisa foi feita com setenta prefeitos e não foi desenhada para ser representativa. Deve ser interpretada como um estudo de caso.

2 EXPERIMENTO DE ADOÇÃO DE POLÍTICA

Nesta seção, descrevemos um experimento de campo que foi desenhado para analisar se o fornecimento de evidências empíricas sobre a efetividade de políticas públicas aos chefes de governos municipais influencia as políticas que são adotadas em seus municípios. Notamos que informar os prefeitos brasileiros sobre a efetividade de uma política que tem como objetivo aumentar a conformidade tributária afeta, causalmente, não apenas as crenças em relação à efetividade da política, mas também a própria adoção da política um a dois anos após o experimento. Esclarecemos a seguir, em detalhes, a política utilizada, assim como o experimento implementado e seus resultados.

2.1 Política: cartas-lembrete ao contribuinte

A essência do experimento de adoção de política é informar a um grupo de prefeitos os resultados de um conjunto de avaliações rigorosas de impacto, evidenciando que cartas-lembrete aumentam a probabilidade de que os contribuintes paguem seus tributos em dia.

Escolhemos a política de cartas-lembrete por três motivos. Em primeiro lugar, o desenho, a aplicação e a arrecadação de impostos estão majoritariamente sob o controle e a responsabilidade dos governos municipais. Porém, como na maioria

dos países em desenvolvimento, a conformidade do contribuinte é um desafio no Brasil. De acordo com De Cesare e Smolka (2004), estima-se, por exemplo, que pelo menos 20% dos contribuintes não cumprem com o pagamento do Imposto Predial e Territorial Urbano (IPTU). Em nossas pesquisas descritivas e grupos focais, os prefeitos participantes relataram um interesse considerável em receber informações sobre como aumentar as receitas tributárias de seus municípios.

Em segundo lugar, a efetividade das cartas-lembrere foi amplamente estudada em diversos experimentos, incluindo dois na América Latina (Coleman, 1996; Hasseldine *et al.*, 2007; Del Carpio, 2013; Fellner, Sausgruber e Traxler, 2013; Castro e Scartascini, 2015; Hallsworth *et al.*, 2017). Intervenções desse tipo mostraram-se surpreendentemente efetivas. Por exemplo, Del Carpio (2013) encontrou que cartas-lembrere incluindo unicamente a data de vencimento dos tributos aumentaram o pagamento de impostos no Peru em 10%, enquanto cartas que também incluíam normas sociais, enfatizando que a maioria das pessoas paga seus impostos no prazo, aumentaram o cumprimento em 20%.

Em terceiro, políticas de cartas-lembrere são relativamente baratas e de fácil implementação, ao mesmo tempo que não possuem um alto custo do ponto de vista político. Por um lado, isso significa que, caso o formulador de políticas seja convencido da efetividade da política apresentada, a probabilidade de que ele venha implementá-la é alta. Por outro lado, cartas-lembrere são provavelmente uma ferramenta de política pública custo-efetiva, em parte porque são de baixo custo e fáceis de implementar.

Cartas-lembrere para os contribuintes são incomuns, mas estão longe de serem inéditas no Brasil. No questionário final, implementado neste experimento, 32% dos municípios do grupo controle relataram o uso de alguma forma de mensagens-lembrere para os contribuintes, envolvendo não somente o envio de cartas, mas também incluindo outros canais de comunicação, como mensagens de texto, folhetos e propaganda na mídia.

2.2 Contexto experimental

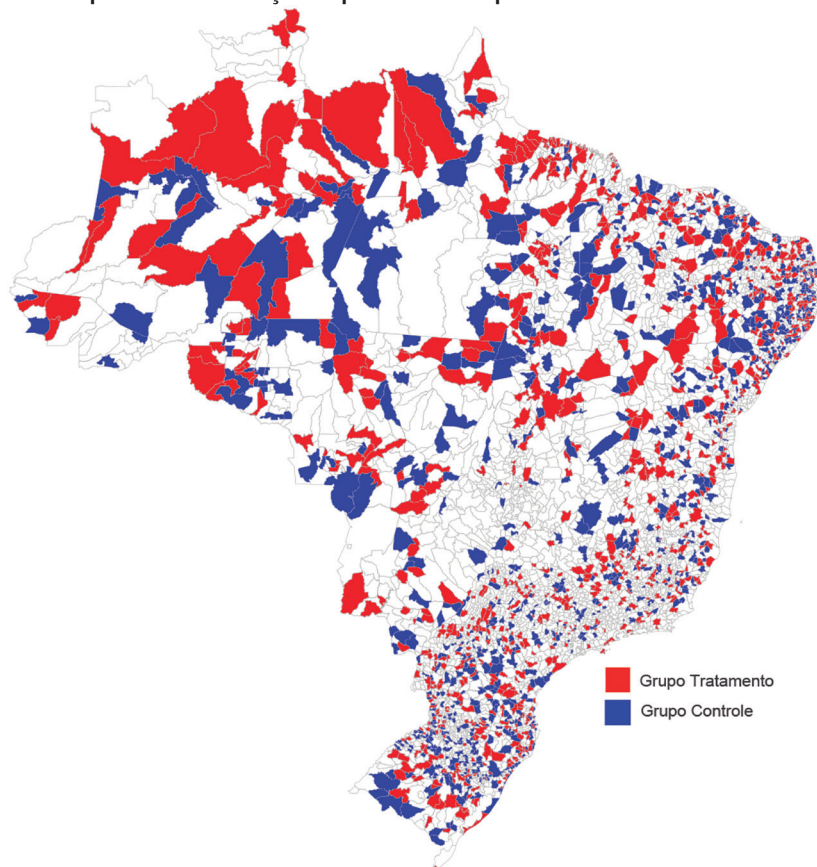
O experimento de adoção de política foi implementado em uma das maiores conferências da CNM: o seminário Novos Gestores, no qual participam prefeitos recém-eleitos e reeleitos. O seminário é realizado a cada quatro anos, geralmente em outubro e novembro, e tem como propósito fornecer formação técnica aos prefeitos que estão prestes a iniciar seus mandatos em janeiro. Os prefeitos podem participar de uma variedade de sessões de treinamento lideradas por especialistas da CNM em tópicos que vão desde planejamento financeiro e orçamentário até áreas de política pública, como desenvolvimento urbano, educação, saúde e turismo. Várias sessões acontecem em paralelo durante a conferência, exceto por um número

limitado de sessões plenárias. A conferência em si decorre em etapas, com prefeitos de diferentes regiões do país presentes em dias diferentes, devido às limitações de capacidade. Cada prefeito comparece por dois dias.

A amostra do experimento é constituída por prefeitos dos municípios com populações entre 5 mil e 100 mil habitantes e que participaram do seminário *Novos Gestores*, realizado entre outubro e novembro de 2016. A amostra total é composta por 1.818 municípios, o que representa 45% de todos os municípios brasileiros nessa faixa populacional. A figura 1 ilustra a distribuição geográfica dos municípios da amostra.

FIGURA 1

Experimento de adoção de política: municípios da amostra



Elaboração dos autores.

A tabela 1 fornece estatísticas descritivas sobre a amostra de prefeitos e municípios. Vemos, por exemplo, que quase 90% dos prefeitos são homens, cerca de

60% possuem pelo menos um diploma de bacharel e 16% estão em seu segundo e último mandato.⁶ O município médio da amostra tem uma população de cerca de 21 mil habitantes.

TABELA 1
Experimento de adoção de política: estatísticas descritivas e balanceamento

Variáveis	Linha de base final			Linha final		
	Média controle (%)	Δ tratamento	P-valor	Média controle (%)	Δ tratamento	P-valor
Características dos prefeitos						
Homem	88,26	1,41	0,34	90,01	-0,14	0,93
Idade	46,76	1,32	0,01	47,08	1,61	0,00
Ensino superior ou mais	57,74	-0,76	0,74	57,66	0,73	0,78
Segundo mandato	15,69	1,56	0,37	15,18	0,91	0,63
Margem de vitória eleitoral	16,73	0,36	0,68	16,61	0,46	0,63
Partido político de esquerda	32,98	2,10	0,35	32,76	1,36	0,58
Características dos municípios						
População	20,86	-0,06	0,94	20,23	0,06	0,95
População com ensino superior	5,17	-0,15	0,25	5,47	-0,14	0,31
Funcionário público com ensino superior	32,50	0,89	0,21	33,32	0,25	0,74
Pobreza	26,40	-0,27	0,76	23,05	0,11	0,91
Gini	50,33	-0,19	0,54	49,37	0,17	0,61
Grande sul	51,01	-0,62	0,79	59,92	-2,36	0,36
Renda <i>per capita</i>	457,64	3,42	0,75	489,23	2,78	0,81
Receitas de impostos municipais (2010-2015)	6,06	0,09	0,68	6,40	0,08	0,75
F-test conjunto	-	-	0,17	-	-	0,20
Atrito						
Município	-	-	-	19,85	-1,69	0,36
Prefeito	-	-	-	48,35	2,28	0,33
Equipe de finanças	-	-	-	24,97	-0,80	0,69

Elaboração dos autores.

Obs.: Médias amostrais do grupo de controle e diferenças de médias em relação ao grupo de tratamento na linha de base e na linha final. Havia 937 (751) municípios no grupo de controle e 881 (721) no grupo de tratamento na linha de base (na linha final). O primeiro bloco de variáveis relata características dos prefeitos que administram os municípios. Partido político de esquerda equivale a 1 para prefeitos pertencentes a um partido de centro-esquerda, de acordo com plataformas políticas históricas, e 0 para caso contrário. O segundo bloco de variáveis reporta as características do município. População é o número de habitantes do município (em milhares). População com ensino superior é a parcela de adultos do município com formação superior. Funcionário público com ensino superior é a parcela de funcionários públicos municipais com formação superior. Pobreza é a taxa de pobreza do município. Gini é o coeficiente de Gini do município. Grande sul é igual a 1 para os municípios das regiões Sul, Sudeste e Centro-Oeste, 0 para caso contrário. A renda *per capita* é a renda mensal *per capita* do município. Receitas de impostos municipais (2010-2015) indica a participação média das receitas tributárias locais no total das receitas municipais de 2010 a 2015. Teste F de significância conjunta e taxa de atrito na linha final do experimento – município, prefeito e equipe de finanças.

6. Esta baixa participação de prefeitos em seu segundo mandato é explicada, em parte, pela crise política que o Brasil estava passando na época das eleições municipais de 2016, o que levou a uma diminuição na proporção de políticos em exercício vencendo a reeleição.

2.3 Desenho experimental

A amostra de prefeitos foi dividida aleatoriamente em grupos de tratamento (n = 881) e controle (n = 937).⁷ Todos os prefeitos podiam comparecer livremente a quaisquer das sessões regulares de treinamento do Novos Gestores, mas apenas prefeitos do grupo de tratamento foram convidados, por *e-mail* e mensagem de texto, para participarem de nossas sessões informativas sobre pesquisas de avaliações de impacto – intervenção experimental. A nossa sessão foi anunciada com o tema *como aumentar as receitas tributárias locais* e enquadrada como uma sessão de treinamento organizada pela CNM, em conjunto com pesquisadores das universidades de Columbia e Harvard. Como a participação era opcional, nosso experimento deve ser considerado como tendo um desenho de promoção aleatória (*encouragement design*). A tabela 1 mostra que os grupos de tratamento e controle estão amplamente balanceados nas características dos prefeitos, bem como nas características municipais.

As sessões de informação duraram 45 minutos e foram conduzidas por um experiente instrutor brasileiro, sem a presença de investigadores estrangeiros. O instrutor começou a sessão apresentando e definindo o que é o impacto de uma política pública, a relação custo-benefício de uma política, e o que são pesquisas de avaliação de impacto. Em seguida, forneceu uma descrição das cartas-lembrete ao contribuinte, incluindo uma carta-exemplo. Finalmente, apresentou os resultados (ou seja, o tamanho dos efeitos estimados) de um conjunto de estudos rigorosos que avaliaram o impacto de cartas-lembrete aos contribuintes. A apresentação enfatizou as características que uma carta-lembrete deveria possuir (segundo as evidências) para ser efetiva, por exemplo, indicar o prazo de pagamento do imposto, mencionar a possibilidade de multas e auditorias pelo não pagamento dos tributos em dia e afirmar que a maioria das pessoas paga seus impostos em dia, assim como enfatizou os efeitos estimados pelos estudos científicos.

As informações apresentadas foram simplificadas, e a apresentação foi concisa. Evitaram-se jargões acadêmicos e tabelas de regressão. A apresentação de trinta minutos foi seguida por uma sessão de quinze minutos aberta para perguntas dos prefeitos.⁸ No final da sessão, os prefeitos receberam um documento resumido

7. A aleatorização foi estratificada com base: no nível de escolaridade do prefeito, se o prefeito estava no segundo mandato, no nível médio de educação entre os funcionários públicos municipais, no tamanho da população do município, no coeficiente de Gini e na região geográfica. Uma parcela um pouco maior de municípios foi designada para o grupo de controle devido a questões logísticas associadas à nossa capacidade de gerenciar um grande número de participantes do grupo de tratamento e à capacidade da sala que a CNM designou para o experimento.

8. Durante os quinze minutos reservados para perguntas, os prefeitos fizeram questionamentos interessantes sobre cartas-lembrete e outras políticas alternativas sobre conformidade tributária: por exemplo, se os efeitos seriam os mesmos se as mensagens fossem enviadas por *e-mail* ou mensagens de texto, em vez de enviadas por cartas em papel; se a política podia ser usada para encorajar os devedores de impostos a pagar seus débitos; e se incentivos financeiros, como descontos ou loterias para pagamento dos impostos em dia, são políticas efetivas. Evitamos fornecer respostas claras a essas perguntas.

(*policy brief*) produzido profissionalmente com o mesmo conteúdo informativo da apresentação, incluindo referências aos estudos citados.

A sessão de informação ocorreu três ou quatro vezes durante cada etapa da conferência. Os prefeitos do grupo de tratamento, portanto, podiam escolher participar quando seu custo de oportunidade de tempo fosse o mais baixo possível. Metade de nossas sessões colidiu com uma sessão plenária que lecionou aos prefeitos sobre finanças e orçamentos municipais e enfatizou o planejamento financeiro e a responsabilidade fiscal. A outra metade entrou em conflito com horários de expediente da CNM e com uma variedade de sessões paralelas. A nossa sessão não entrou em conflito com outras sessões que tratassem de políticas sociais, políticas de desenvolvimento urbano e políticas econômicas; e nenhuma outra sessão da conferência enfatizou informações de estudos acadêmicos ou avaliações de impacto. A julgar por nossas observações informais, a alternativa mais comum em relação à nossa sessão de informação foi fazer *networking* com outros prefeitos.

2.4 Dados

Para medir se o fornecimento de informações sobre estudos de avaliações de impacto afetou as crenças dos líderes políticos e a adoção de políticas, conduzimos um questionário 15 a 24 meses após a sessão de informação, por telefone, com funcionários relevantes da administração pública dos municípios de tratamento e controle. Procuramos entrar em contato com o gestor responsável pela execução da política tributária de cada município e com o próprio prefeito.⁹

A aplicação do questionário foi supervisionada por um assistente de pesquisa e conduzida por uma equipe de nove entrevistadores que desconheciam a atribuição dos municípios aos grupos de tratamento e controle e as hipóteses do experimento. Após dez meses de telefonemas, entrevistamos com sucesso pelo menos uma pessoa em 81% de nossa amostra de municípios – 75% dos gestores tributários e 51% dos prefeitos da amostra. Não conseguimos fazer nenhum tipo de contato com 10% dos municípios da amostra, por não termos conseguido localizar um número de telefone em funcionamento.¹⁰ Não houve atrito diferencial entre os grupos de tratamento e controle, e as características observáveis dos municípios contatados com sucesso são semelhantes em ambos os grupos, conforme relatado na tabela 1.

9. Normalmente, os secretários de finanças são os responsáveis pela divisão tributária dos municípios brasileiros. No entanto, solicitamos especificamente aos atendentes do telefonema que passassem a chamada para o responsável pela divisão de impostos. Assim que fomos transferidos, confirmamos se a pessoa realmente ocupava aquele cargo ou solicitamos o número de telefone do responsável pela administração da política tributária municipal.

10. Em média, muitas horas de trabalho foram necessárias para que pudéssemos falar com os gestores tributários e prefeitos, principalmente para coletar os números de telefone dos municípios. Nem todos os municípios brasileiros publicam ou têm informações de contato atualizadas em seus *sites*, então coletamos números de telefone por meio de pesquisas no Google e no Facebook e ligamos para outras instituições locais, como hospitais e escolas.

O questionário durou aproximadamente quinze minutos. Indagamos se o município enviou aos contribuintes lembretes para o pagamento de seus impostos e se as mensagens enviadas apresentavam as características descritas em nossas sessões informativas, como: o prazo de vencimento dos tributos, a possibilidade de multas ou auditorias e a norma social de pagar os impostos em dia. Os participantes também foram questionados sobre suas crenças em relação à efetividade da política de cartas-lembrete. Especificamente, perguntamos sobre o provável impacto de tal política em seus municípios, em termos de mudanças percentuais na conformidade tributária, mesmo se o município relatou não usar cartas-lembrete. Além disso, fizemos perguntas que serviram como verificações de atenção e compreensão, bem como perguntas sobre um possível substituto da política de cartas-lembrete (incentivos financeiros para contribuintes) e perguntas placebo nas quais esperaríamos efeitos do tratamento nulos (o uso de pregão eletrônico nas compras da prefeitura).

Além dos dados coletados nesse questionário final, coletamos dados demográficos, eleitorais e orçamentários de fontes oficiais para todos aqueles municípios que tinham dados disponíveis.¹¹ Como não é possível observar a conformidade tributária nos dados oficiais, nossa variável de interesse principal é se o município implementou a política de lembretes tributários. Levando em consideração que possam existir vieses de relato e efeitos de demanda nas respostas dos prefeitos, na próxima subseção, discutimos separadamente as respostas dos prefeitos e dos gestores tributários.

2.5 Resultados

2.5.1 Participação nas sessões de informação

Optaram por participar de nossa sessão 37,9% dos prefeitos do grupo de tratamento. Em contraste, menos de 1% dos prefeitos do grupo de controle compareceu à sessão. Os custos de oportunidade de participar – renunciar à oportunidade de participar de outras sessões de treinamento paralelas, ou conduzir reuniões com outros prefeitos, políticos e funcionários – foram significativos, embora difíceis de quantificar. Além disso, alguns prefeitos não tinham informações de contato precisas armazenadas no sistema da CNM e, portanto, não receberam nossas mensagens de convite. Em vista disso, consideramos 37,9% uma taxa de participação bastante alta.

11. Os dados demográficos são disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e podem ser encontrados em: <<https://bit.ly/3ewUhi3>> e <<https://bit.ly/3EEZHlo>>. O Tribunal Superior Eleitoral do Brasil fornece dados sobre resultados eleitorais e características dos prefeitos (disponível em: <<https://bit.ly/3HZQP5Q>>). Os dados sobre a escolaridade da administração pública foram obtidos da Relação Anual de Informações Sociais (Raís), que é coletada e compilada anualmente pelo Ministério do Trabalho (disponível em: <<https://bit.ly/3pCbmGN>>). Os dados orçamentários foram obtidos do Tesouro Nacional, que compila e divulga anualmente registros contábeis autorrelatados de todos os municípios brasileiros (disponíveis em: <<https://bit.ly/34EB7Vy>> e <<https://bit.ly/3FyhYlq>>).

A análise de preditores de participação na sessão de informação, relatada na tabela 2, revela que prefeitos mais jovens e com educação universitária são, respectivamente, 7 pontos percentuais (p.p.) e 15 p.p. mais prováveis de participarem do que outros, mas, em contrapartida, prefeitos no segundo mandato não têm menor probabilidade de participar do que prefeitos em seu primeiro mandato. Nenhuma das características municipais, como taxas de pobreza, desigualdade ou renda *per capita*, prediz a participação no experimento.

TABELA 2
Experimento de adoção de política: preditores individuais e municipais de participação na sessão de informação

Variável	(1) Sessão de informação	(2) Sessão de informação	(3) Sessão de informação
Características dos prefeitos			
Homem	0,0157 (0,0546)	-	-0,0014 (0,0560)
Idade	-0,0719 (0,0328)	-	-0,0771 (0,0336)
Ensino superior ou mais	0,1616 (0,0328)	-	0,1562 (0,0333)
Segundo mandato	-0,0007 (0,0448)	-	0,0057 (0,0456)
Margem de vitória eleitoral	0,0265 (0,0326)	-	0,0231 (0,0330)
Partido político de esquerda	0,0314 (0,0347)	-	0,0379 (0,0352)
Características dos municípios			
População	-	-0,0079 (0,0343)	-0,0141 (0,0340)
População com ensino superior	-	0,0634 (0,0466)	0,0492 (0,0463)
Funcionário público com ensino superior	-	-0,0345 (0,0339)	-0,0442 (0,0337)
Pobreza	-	-0,1015 (0,0903)	-0,0753 (0,0926)
Gini	-	0,0449 (0,0382)	0,0462 (0,0382)
Grande sul	-	0,0258 (0,0662)	0,0593 (0,0660)
Renda <i>per capita</i>	-	-0,0762 (0,0839)	-0,0663 (0,0854)
Receitas de impostos municipais (2010-2015)	-	-0,0245 (0,0459)	-0,0166 (0,0454)
Constante	0,2876 (0,0631)	0,4343 (0,0937)	0,3376 (0,1124)
Observações	874	878	871

Elaboração dos autores.

Obs.: Estimativa de mínimos quadrados ordinários (MQO). A variável dependente é uma *dummy* que é igual a 1 se o prefeito participou da sessão de informação, e 0 para caso contrário. O modelo é estimado considerando unicamente a amostra de municípios do grupo de tratamento. Expressamos todas as variáveis contínuas como indicadores de acima-abaixo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0, prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas). As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, em que 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); e participação das receitas tributárias locais acima e abaixo da mediana (1/0). Erros-padrão robustos agrupados no nível do município estão entre parênteses.

2.5.2 Adoção de política

Observamos que a sessão de informação sobre resultados de avaliações de impacto leva a um aumento entre 10 p.p. e 11 p.p. no uso de cartas-lembrete aos contribuintes no grupo de tratamento. Isso é equivalente a um aumento de aproximadamente 33% em relação à parcela dos municípios do grupo de controle que começaram a usar essa política de cartas-lembrete em algum momento durante o período analisado.

A tabela 3 apresenta as estimativas do efeito do tratamento sobre os tratados (ToT), usando a atribuição aleatória ao tratamento como um instrumento para a participação na sessão de informação (resultados também apresentados no gráfico 2). A variável dependente é uma *dummy* igual a 1 se o respondente relatou que a política de lembretes tributários é usada em seu município, e 0 caso contrário. Os erros-padrão são agrupados no nível municipal. Na coluna 1, o coeficiente ToT é de 10,3 p.p. (erros-padrão 5,3 p.p.), em comparação com 31,7 p.p. do grupo de controle. Adicionar controles na coluna 2 deixa a estimativa praticamente inalterada. A coluna 3 elimina os participantes que falharam na pergunta que tinha como objetivo verificar a atenção do respondente às perguntas do questionário por telefone, deixando novamente o coeficiente inalterado.¹² Mais importante, as estimativas são muito semelhantes se restringirmos nossa atenção às respostas de prefeitos (coluna 4) ou gestores tributários (coluna 5). Diante disso, e tendo em vista que os funcionários dos departamentos tributários não participaram de nossas sessões de informação, aumenta nossa confiança de que os efeitos estimados não são causados por vieses de resposta.

12. A pergunta de verificação de atenção foi: "Os lembretes tributários enviados aos contribuintes os informaram de que a constituição brasileira foi reformada em 1988". Uma vez que consideramos isso extremamente improvável como texto para um lembrete tributário, inferimos que os participantes que responderam "sim" a esta pergunta simplesmente não estavam prestando atenção ou compreendendo as perguntas.

TABELA 3
Experimento de adoção de política: lembretes tributários

Variável	(1) Adotou	(2) Adotou	(3) Adotou	(4) Adotou	(5) Adotou
Sessão de informação	0,1031 (0,0531)	0,1065 (0,0526)	0,1024 (0,0546)	0,1177 (0,0791)	0,1094 (0,0653)
Observações	2.271	2.239	2.027	898	1.341
Respondentes	Todos	Todos	Todos	Prefeitos	Gestores tributários
Sem desatentos	Não	Não	Sim	Não	Não
Características dos prefeitos	Não	Sim	Sim	Sim	Sim
Características dos municípios	Não	Sim	Sim	Sim	Sim
Municípios	1.465	1.447	1.395	898	1.341
Média do grupo controle	0,317	0,314	0,294	0,364	0,280

Elaboração dos autores.

Obs.: Estimativas de mínimos quadrados em dois estágios (MQ2E). A variável dependente é uma *dummy* que é igual a 1 se o respondente diz que a política foi adotada no município, e 0 para caso contrário. Sessão de informação é uma *dummy* que é igual a 1 se o prefeito do município participou da sessão de informação sobre cartas-lembrete aos contribuintes, 0 para caso contrário. Esta última variável é instrumentalizada com a atribuição ao tratamento. Nas linhas abaixo dos coeficientes, sem desatentos refere-se a se os respondentes que responderam incorretamente à pergunta de verificação de atenção ao questionário de linha final são excluídos do modelo, no qual a pergunta de verificação de atenção foi "Os lembretes tributários enviados aos contribuintes os informaram de que a constituição brasileira foi reformada em 1988". Expressamos todas as variáveis contínuas como indicadores de acima-abaxo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0), prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas). As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, onde 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); e participação das receitas tributárias locais acima e abaixo da mediana (1/0). Erros-padrão robustos agrupados no nível do município estão entre parênteses.

A análise exploratória de heterogeneidades nos efeitos do tratamento por características dos prefeitos e dos municípios (tabela de resultados não reportada neste capítulo) não revela nenhuma evidência clara de efeitos heterogêneos, em parte, talvez, pelo limitado poder estatístico da análise. Os sinais dos coeficientes estimados sugerem que prefeitos no segundo mandato não são substancialmente menos propensos a adotar cartas-lembrete, prefeitos com idade e margens de vitória eleitoral acima da mediana parecem menos propensos a adotar cartas-lembrete, e prefeitos de esquerda parecem mais propensos a adotar esse tipo de política. Porém, é importante frisar que nenhuma dessas estimativas é estatisticamente significativa.

A tabela 4, a seguir, relata os efeitos do tratamento separadamente para os diferentes tipos de mensagens que uma carta-lembrete poderia incluir, segundo as evidências apresentadas na nossa sessão de informação. Os efeitos são bastante semelhantes na probabilidade de usar cartas que: enfatizam a data de vencimento dos tributos; especificam a possibilidade de auditorias e/ou penalidades por não pagar os tributos em dia; e mencionam as normas sociais de pagar os impostos –

embora este último tipo de mensagem tenha um efeito maior em termos relativos, uma vez que é particularmente improvável que seja usado nos municípios do grupo de controle.

TABELA 4
Experimento de adoção de política: características dos lembretes tributários

Variável	(1) Data de vencimento	(2) Auditorias	(3) Norma social
Sessão de informação	0,1014 (0,0522)	0,0720 (0,0471)	0,0990 (0,0374)
Observações	2.239	2.239	2.239
Respondentes	Todos	Todos	Todos
Sem desatentos	Não	Não	Não
Características dos prefeitos	Sim	Sim	Sim
Características dos municípios	Sim	Sim	Sim
Municípios	1.447	1.447	1.447
Média do grupo controle	0,310	0,220	0,112

Elaboração dos autores.

Obs.: Estimativas de MQ2E. A variável dependente é uma *dummy* que é igual a 1 se o respondente diz que a política foi adotada no município, 0 para caso contrário. Data de vencimento refere-se a uma mensagem de lembrete destacando o prazo de pagamento do imposto. Auditoria refere-se a uma mensagem de lembrete destacando a possibilidade de auditorias por não pagar o imposto em dia. Norma social refere-se a uma mensagem de lembrete destacando a norma social de pagamento de impostos. Sessão de informação é uma *dummy* que é igual a 1 se o prefeito do município participou da sessão de informação sobre cartas-lembrete aos contribuintes, 0 para caso contrário. Esta última variável é instrumentalizada com a atribuição ao tratamento. Nas linhas abaixo dos coeficientes, sem desatentos refere-se a se os respondentes que responderam incorretamente à pergunta de verificação de atenção ao questionário de linha final são excluídos do modelo, no qual a pergunta de verificação de atenção foi "Os lembretes tributários enviados aos contribuintes os informaram de que a constituição brasileira foi reformada em 1988". Expressamos todas as variáveis contínuas como indicadores de acima-abaixo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0, prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas). As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, onde 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); e participação das receitas tributárias locais acima e abaixo da mediana (1/0). Erros-padrão robustos agrupados no nível do município estão entre parênteses.

Finalmente, a tabela 5 relata efeitos do tratamento nulos sobre o uso de preção eletrônico nas compras do governo municipal – pergunta placebo – e sobre o uso de incentivos financeiros para o cumprimento de impostos – uma política relativamente comum nos municípios brasileiros, que concebivelmente poderia ter sido considerada como substituta da política de cartas-lembrete.

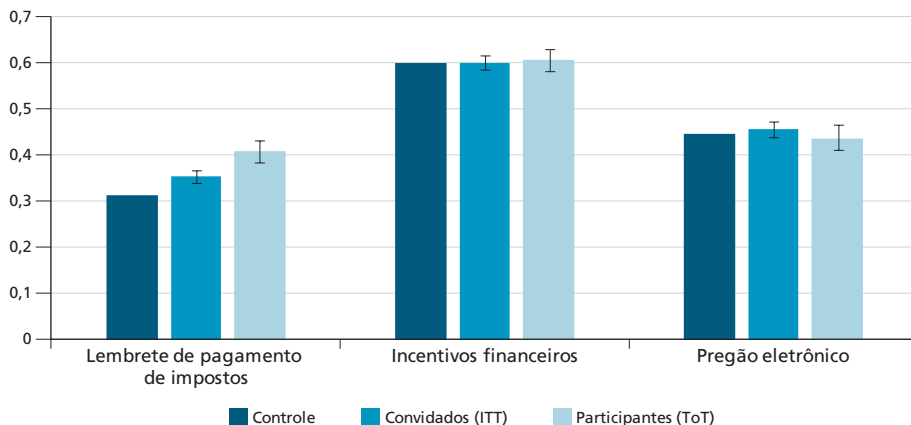
TABELA 5
Experimento de adoção de política: incentivos financeiros e pregão eletrônico

Variável	(1) Incentivos financeiros	(2) Pregão eletrônico
Sessão de informação	0,0033 (0,0557)	0,0153 (0,0644)
Observações	2.177	1.675
Respondentes	Todos	Todos
Sem desatentos	Não	Não
Características dos prefeitos	Sim	Sim
Características dos municípios	Sim	Sim
Municípios	1.434	1.178
Média do grupo controle	0,600	0,447

Elaboração dos autores.

Obs.: Estimativas de MQ2E. A variável dependente é uma *dummy* que é igual a 1 se o respondente diz que a política foi adotada no município, 0 para caso contrário. Sessão de informação é uma *dummy* que é igual a 1 se o prefeito do município participou da sessão de informação sobre cartas-lembrete aos contribuintes, 0 para caso contrário. Esta última variável é instrumentalizada com a atribuição ao tratamento. Nas linhas abaixo dos coeficientes, sem desatentos refere-se a se os respondentes que responderam incorretamente à pergunta de verificação de atenção ao questionário de linha final são excluídos do modelo, no qual a pergunta de verificação de atenção foi “Os lembretes tributários enviados aos contribuintes os informaram de que a constituição brasileira foi reformada em 1988”. Expressamos todas as variáveis contínuas como indicadores de acima-abaxo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0), prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas. As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, onde 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); e participação das receitas tributárias locais acima e abaixo da mediana (1/0). Erros-padrão robustos agrupados no nível do município estão entre parênteses.

GRÁFICO 2
Adoção de política: resultado principal



Elaboração dos autores.

Obs.: 1. Proporção de participantes por grupo que disseram que a respectiva política foi adotada em seus municípios. As políticas indagadas foram: o envio aos contribuintes de lembretes tributários, incentivos financeiros para os contribuintes pagarem seus tributos e o uso de pregão eletrônico nas compras da prefeitura. Controle refere-se ao grupo que não recebeu convite; convidados (ITT), ao grupo que recebeu convite para participar da sessão informativa; e participantes (ToT), grupo que recebeu o convite e compareceu à sessão.

2. Em torno das estimativas pontuais, relatam-se intervalos de confiança de 95% de erros-padrão robustos agrupados no nível do município.

2.5.3 Crenças

Por meio do questionário final, também medimos as crenças dos participantes sobre a efetividade da política de cartas-lembrere. Crenças sobre a efetividade da política é um mecanismo plausível por meio do qual o impacto da sessão informativa sobre a adoção de cartas-lembrere poderia operar. Perguntamos aos prefeitos e aos gestores tributários qual seria o efeito da política caso fosse implementada em seu município, independentemente se a política foi de fato implementada ou não. Comparamos suas crenças com um efeito estimado de 12%, que foi a principal estimativa de cartas-lembrere que compartilhamos com os prefeitos durante as sessões de informação.

O painel A da tabela 6 mostra que comparecer à sessão de informação – instrumentalizada usando a atribuição aleatória ao tratamento – aumentou a “precisão” das crenças, mesmo 15 a 24 meses após o tratamento. Especificamente, o desvio absoluto das crenças no grupo de tratamento em relação ao tamanho do efeito mencionado na sessão de informação é 20% menor do que no grupo de controle. Comparando as colunas 4 e 5, podemos ver que as crenças se tornaram mais precisas não apenas entre os prefeitos, mas também entre os gestores tributários, o que sugere a existência de um fluxo de informação dentro do governo municipal. Isso talvez tenha sido facilitado ao fornecer aos prefeitos participantes um documento compartilhável contendo o resumo da nossa sessão.

O painel B da tabela 6 estima o efeito da precisão das crenças na adoção de cartas-lembrete, agora instrumentalizando a precisão das crenças com a atribuição aleatória ao tratamento. Certamente, isso requer fazer a discutível suposição de que o tratamento afeta a adoção da política apenas por meio das crenças. As estimativas implicam que aumentar a precisão da crença em 1 p.p. (ou seja, reduzir o desvio absoluto em 1 p.p. em uma base de cerca de 7 p.p.) aumenta a adoção da política em 8 p.p. (erros-padrão 5 p.p.). Obviamente, os efeitos podem operar também por meio de outros canais, como confiança, relevância etc., conforme discutido em seguida. Esses achados devem, portanto, ser tratados como puramente descritivos. Também é importante notar que as crenças relevantes eram presumivelmente aquelas ponderadas no momento da decisão de adoção da política, informação que não temos disponível.

TABELA 6
Experimento de adoção de política e precisão das crenças: lembretes tributários

Painel A	(1)	(2)	(3)	(4)	(5)
Variável	Precisão das crenças	Precisão das crenças	Precisão das crenças	Precisão das crenças	Precisão das crenças
Sessão de informação	1,3975 (0,5209)	1,3541 (0,5201)	1,5031 (0,5589)	1,1923 (0,7396)	1,5125 (0,6839)
Média do grupo controle	-6,980	-6,983	-6,998	-6,869	-7,060
Painel B	(1)	(2)	(3)	(4)	(5)
Variável	Adotou	Adotou	Adotou	Adotou	Adotou
Precisão das crenças	0,0856 (0,0500)	0,0935 (0,0537)	0,0819 (0,0483)	0,1344 (0,1084)	0,0799 (0,0562)
Média do grupo controle	0,310	0,306	0,285	0,357	0,271
Observações	2.172	2.141	1.936	842	1.299
Respondentes	Todos	Todos	Todos	Prefeitos	Gestores tributários
Sem desatentos	Não	Não	Sim	Não	Não
Características dos prefeitos	Não	Sim	Sim	Sim	Sim
Características dos municípios	Não	Sim	Sim	Sim	Sim
Municípios	1.434	1.416	1.360	842	1.299

Elaboração dos autores.

Obs.: Estimativas de MQZE. Atribuição de tratamento é o instrumento para sessão de informação (painel A) e para precisão das crenças (painel B). No painel A, a variável dependente – precisão das crenças – é a diferença absoluta multiplicada por -1 entre as crenças dos participantes em relação ao efeito que os lembretes tributários teriam sobre as receitas tributárias locais e o efeito de 12% informado durante a sessão de informação. Sessão de informação é uma *dummy* que é igual a 1 se o prefeito do município participou da sessão de informação sobre cartas-lembrete aos contribuintes, 0 para caso contrário. No painel B, a variável dependente é uma *dummy* que é igual a 1 se o respondente diz que a política foi adotada no município, 0 para caso contrário. Precisão das crenças é a diferença absoluta multiplicada por -1 entre as crenças dos participantes em relação ao efeito que os lembretes tributários teriam sobre as receitas tributárias locais e o efeito de 12% informado durante a sessão de informação. Nas linhas abaixo dos coeficientes, sem desatentos refere-se a se os respondentes que responderam incorretamente à pergunta de verificação de atenção ao questionário de linha final são excluídos do modelo, no qual a pergunta de verificação de atenção foi “Os lembretes tributários enviados aos contribuintes os informaram de que a constituição brasileira foi reformada em 1988”. Expressamos todas as variáveis contínuas como indicadores de acima-abaixo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0, prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas). As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, onde 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); e participação das receitas tributárias locais acima e abaixo da mediana (1/0). Erros-padrão robustos agrupados no nível do município estão entre parênteses.

2.6 Experimento de adoção de política: discussão

Este experimento tem um resultado simples, mas importante: quando os líderes políticos no Brasil recebem informações de estudos acadêmicos sobre o impacto de uma política custo-efetiva que tem como objetivo aumentar a conformidade tributária, eles mudam as políticas em uso em suas jurisdições. Isso implica que os formuladores de políticas públicas estão abertos a novas evidências, preocupam-se com a efetividade das políticas e possuem pelo menos alguma capacidade e desejo de traduzir evidências científicas em decisões práticas de política.

Algumas ressalvas são importantes de serem discutidas. Em primeiro lugar, não podemos descartar que os efeitos estimados sejam explicados, em parte, pelo simples fato de os prefeitos ficarem sabendo da existência da política de cartas-lembrete, e não devido aos seus impactos informados pelos estudos científicos apresentados. Conforme observado nos resultados, no entanto, a política de cartas-lembrete ao contribuinte está longe de ser desconhecida no Brasil. Aproximadamente um terço dos municípios do grupo de controle relatou usar alguma forma de lembrete tributário. Em contrapartida, também encontramos evidência de que as crenças são mais precisas no grupo de tratamento, embora não possamos descartar que os efeitos teriam sido semelhantes se simplesmente tivéssemos fornecido uma recomendação de política sem qualquer tipo de evidência científica.

Em segundo lugar, o experimento está baseado em uma política que é barata, relativamente fácil de implementar e que tem um objetivo preciso em um contexto econômico que ampara sua implementação. Outras políticas efetivas podem ter custos de implementação mais elevados, requererem uma maior capacidade técnica para ser implementadas, ou serem mais controversas do ponto de vista político ou ideológico. Nesses casos, a mudança de crenças sobre a efetividade de um programa pode não se traduzir em mudança de políticas adotadas.

Em terceiro lugar, estimamos o efeito de fornecer informações sobre estudos de avaliações de impacto em um contexto particular, no caso uma sessão de informação elaborada e organizada por pesquisadores de universidades estrangeiras de renome, e em uma conferência organizada por uma organização local confiável. Resultados de estudos científicos recebidos de outras fontes, como centros de pesquisas locais, organizações não governamentais (ONGs), ou da mídia, podem ser ponderados de forma diferente. Da mesma forma, os formuladores de políticas públicas que buscam localizar informações relevantes sobre estudos científicos podem ter dificuldade em encontrar e interpretar essas informações. Por sua vez, nosso experimento de adoção de política também não captura os numerosos canais menos diretos por meio dos quais estudos científicos podem influenciar a escolha de políticas, como mudar gradualmente as formas de pensar e influenciar doadores e outros atores estatais e não estatais.

3 EXPERIMENTO DE CRENÇAS

Nesta seção, descrevemos um experimento que foi desenhado para medir: i) se os formuladores de políticas municipais no Brasil possuem interesse em se informar sobre estudos de avaliação de impacto; e ii) como tais informações afetam suas crenças em relação à efetividade de políticas. A área de política em que este experimento se concentrou foi em programas de desenvolvimento na primeira infância (DPI), um tópico muito estudado nas ciências sociais. Observamos que os formuladores de políticas públicas valorizam os resultados de estudos científicos sobre os efeitos de programas de DPI e que atualizam suas crenças substancialmente em resposta às informações dos estudos. A descrição do experimento e de seus resultados é fornecida a seguir.

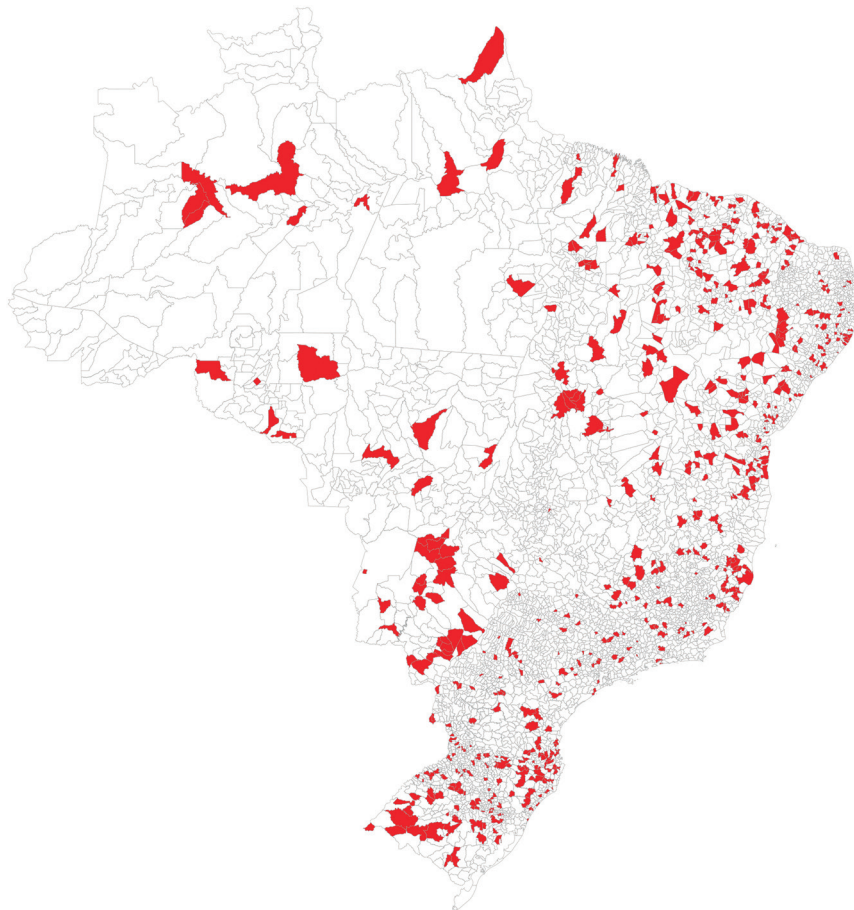
3.1 Contexto experimental

O experimento de crenças foi implementado com 764 funcionários municipais de 579 municípios em catorze conferências da CNM em todo o Brasil, entre 2017 e 2018.¹³ As conferências tiveram a participação de prefeitos, vice-prefeitos, secretários municipais e vereadores. Desenhamos um experimento de meia hora de duração que foi autoadministrado pelos participantes usando *tablets*. O experimento não foi anunciado com antecedência aos participantes. Em vez disso, os assistentes de pesquisa recrutaram participantes das conferências durante os intervalos entre as sessões regulares da CNM, conforme descrito na subseção 3.2.

Quase 50% dos participantes do experimento eram prefeitos; 26%, vereadores; 18%, secretários municipais; e 6%, vice-prefeitos. A distribuição geográfica dos municípios representados pelos participantes é ilustrada na figura 2, e a tabela 7 fornece estatísticas descritivas. Cerca de 38% dos municípios representados têm prefeitos filiados a um partido político de esquerda, e, aproximadamente, 20% das crianças de 0 a 3 anos e 78% das crianças de 4 a 5 anos nestes municípios frequentam a pré-escola. Quarenta e dois por cento dos participantes relataram que seus municípios implementaram programas de DPI.

13. As conferências compreenderam duas conferências nacionais realizadas em Brasília (maio de 2017 e 2018) e doze conferências regionais Diálogo Municipalista organizadas de agosto a dezembro de 2017 em Alagoas, na Bahia, no Ceará, no Espírito Santo, no Maranhão, em Mato Grosso do Sul, em Minas Gerais, no Paraná, no Piauí, no Rio Grande do Sul, em Santa Catarina e em São Paulo. Além disso, outro grupo de 134 funcionários municipais de 117 municípios completou um questionário descritivo sobre as vantagens e desvantagens dos diferentes estudos usados neste experimento.

FIGURA 2
Experimento de crenças: municípios da amostra



Elaboração dos autores.

TABELA 7
Experimento de crenças: estatísticas descritivas e balanceamento

Variáveis	Média controle (%)	Δ desenvolvimento	P-valor	Δ grande	P-valor
Características dos prefeitos					
Homem	91,46	-1,32	0,38	-4,04	0,00
Idade	48,61	-0,11	0,82	-0,49	0,28
Ensino superior	57,62	2,80	0,20	0,13	0,95
Segundo mandato	18,29	0,44	0,82	-0,07	0,97
Margem de vitória eleitoral	14,19	0,18	0,81	0,36	0,55
Partido político de esquerda	38,72	-0,99	0,64	1,80	0,42
Características dos municípios					
População	24,49	1,45	0,48	1,24	0,40
População com ensino superior	4,915	-0,07	0,52	0,02	0,87
Funcionário público com ensino superior	34,16	-0,97	0,09	-0,85	0,17
Pobreza	26,45	0,48	0,55	0,41	0,61
Gini	49,48	0,48	0,09	0,44	0,13
Grande sul	51,22	-0,75	0,74	-1,07	0,63
Renda <i>per capita</i>	457,1	-8,79	0,40	1,02	0,93
Crianças na escola (0-3 anos)	19,88	-1,04	0,04	0,08	0,87
Crianças na escola (4-5 anos)	78,34	-0,41	0,54	0,16	0,83
Características questionário etapa introdutória					
Prefeito	49,70	-0,16	0,94	-1,33	0,55
Político de carreira	29,27	0,74	0,72	-0,47	0,81
Esquerdista	23,78	-2,37	0,19	-1,79	0,35
Implementou DPI	41,77	0,40	0,85	-3,02	0,17
Ouviu sobre DPI	26,22	-0,81	0,68	-0,11	0,95

Elaboração dos autores.

Obs.: Médias amostrais das observações de controle. Δ desenvolvimento e Δ grande relatam o coeficiente estimado, com seu respectivo *p*-valor, de uma regressão linear de cada característica do prefeito, do município e do questionário da etapa introdutória do experimento, em duas variáveis *dummy*. Uma *dummy* que é igual a 1 para Jamaica e Colômbia e 0 para caso contrário (*desenvolvimento*), e uma *dummy* que é igual a 1 para Colômbia e Estados Unidos e 0 para caso contrário (*grande*). Observações de controle são aquelas para as quais a *dummy desenvolvimento* e a *dummy grandes* são iguais a 0. A regressão linear é estimada com 1.368 observações. Erros-padrão robustos são agrupados no nível individual (764 agrupamentos). A média de controle é calculada com 368 observações. O primeiro bloco de variáveis relata características dos prefeitos que administram os municípios. Partido político de esquerda é igual a 1 para prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas, 0 para caso contrário. O segundo bloco de variáveis reporta as características do município. População é o número de habitantes do município (em milhares). População com ensino superior é a parcela de adultos do município com formação superior. Funcionário público com ensino superior é a parcela de funcionários públicos municipais com formação superior. Pobreza é a taxa de pobreza do município. Gini é o coeficiente de Gini do município. Grande sul é igual a 1 para os municípios das regiões Sul, Sudeste e Centro-Oeste, 0 para caso contrário. A renda *per capita* é a renda mensal *per capita* do município. Crianças na escola (0-3 anos) é a proporção de crianças de 0 a 3 anos no município que frequentam a pré-escola. Crianças na escola (4-5 anos) é a proporção de crianças de 4 a 5 anos no município que frequentam a pré-escola. O terceiro bloco de variáveis relata características autorrelatadas pelos participantes do experimento. Prefeito é a proporção de prefeitos entre os participantes do experimento. Político de carreira é igual a 1 se o participante ocupou cargo eletivo no mandato anterior, 0 para caso contrário. Esquerdista é igual a 1 se o participante se autoidentificou como esquerdista (0-4) em uma escala de 0 a 10, 0 para caso contrário. Implementou DPI é igual a 1 se o participante relatou que o município já implementou um programa de DPI, 0 para caso contrário. Ouviu sobre DPI é igual a 1 se o participante relatou que já ouviu falar sobre programas de DPI, 0 para caso contrário.

Recrutamos 38% dos prefeitos presentes, 49% dos vice-prefeitos, 35% dos secretários municipais e 41% dos vereadores. A participação foi limitada pelo número de *tablets* disponíveis e pelo número de intervalos na programação das conferências, mas os participantes também poderiam ter se autosselecionado para o experimento com base no incentivo de participação que oferecemos (bilhetes de um sorteio) ou por seus interesses em políticas educacionais.¹⁴ Este último motivo, potencialmente, poderia enviesar para cima nossas estimativas de demanda por informações de estudos científicos. Não obstante, a única diferença observada entre os prefeitos participantes do experimento e aqueles que não participaram – mas que estavam presentes nas conferências – é que os prefeitos participantes tinham maior probabilidade (7 p.p.) de serem de partidos de esquerda (tabela de resultados não reportada neste capítulo).

3.2 Desenho experimental

A estrutura do experimento, representada graficamente na figura 3, é descrita a seguir. Começamos apresentando a política de DPI. Em seguida, extraímos as crenças iniciais dos participantes sobre a efetividade da política e sua disposição a pagar (DAP) para saber os resultados de um estudo de avaliação de impacto de DPI. Logo, revelamos os achados do estudo e, por fim, extraímos as crenças posteriores dos participantes para avaliar até que ponto os resultados do estudo afetaram suas crenças.

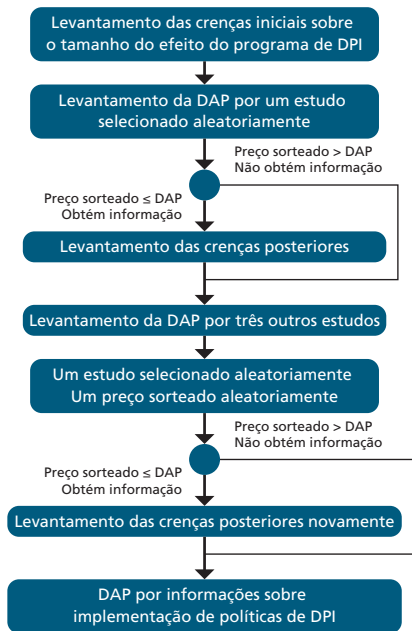
3.2.1 Etapa introdutória

Começamos com um breve questionário levantando informações demográficas e profissionais. Perguntamos aos participantes, por exemplo, sua filiação partidária, suas visões políticas definidas em uma escala esquerda-direita e qual função profissional desempenhava antes do início do mandato 2017-2020 – cargo eletivo, funcionário público, funcionário no setor privado etc. Em seguida, descrevemos as características principais dos programas de DPI: como são geralmente implementados – via creches ou visitas aos domicílios; a frequência e o tipo de atividades que são desenvolvidas – jogos educativos, uso de livros ilustrados, por exemplo, um dia por semana por um período de dois anos; e os objetivos que têm – aumentar as habilidades cognitivas das crianças. Para facilitar a compreensão da política e de seus objetivos, fornecemos referências de programas semelhantes no Brasil, como o programa Criança Feliz. Logo, perguntamos aos funcionários municipais se eles já tinham ouvido falar sobre, ou se os seus municípios tinham implementado, programas com as características descritas.

14. Bilhetes de sorteio são para participar de um evento que debate o Brasil e tem grande visibilidade política. Por vezes, chamamos de bilhete de loteria apenas neste texto para torná-lo menos repetitivo. O termo *bilhete de loteria*, e a conotação que traz, nunca foi usado no trabalho de campo com os gestores.

Finalmente, descrevemos as principais variáveis de resultado por meio das quais esses programas podem ser avaliados – notas de testes – e como esses resultados são relatados – tamanho de efeito padronizado. Novamente, apresentamos exemplos ilustrativos para promover o entendimento dessas variáveis e métricas. Por exemplo, evidenciamos os ganhos em notas de teste associados a um ano adicional de estudo no ensino médio (0,2 desvio-padrão) segundo a Prova Brasil. No final da etapa introdutória, fizemos perguntas de compreensão sobre o que foi apresentado. Os participantes só conseguiam avançar às próximas etapas do experimento após terem respondido corretamente às perguntas de compreensão ou terem recebido auxílio e maiores esclarecimentos por parte da nossa equipe em campo.

FIGURA 3
Experimento de crenças: estrutura



Elaboração dos autores.

3.2.2 Crenças iniciais

Começamos a parte principal do experimento obtendo as crenças iniciais dos participantes. Especificamente, perguntamos o que eles consideravam que seria o efeito da política DPI sobre as habilidades cognitivas das crianças, caso a política fosse implementada em seu próprio município.¹⁵ É importante salientar que as crenças

15. Para simplificar, e devido ao tempo limitado com cada participante, levantamos apenas previsões pontuais (sobre os efeitos nas habilidades cognitivas), em vez de levantar crenças probabilísticas completas. Essa é uma limitação importante do estudo, a qual discutiremos posteriormente.

só podiam ser relatadas em intervalos de 0,1 desvio-padrão entre um mínimo de 0 desvio-padrão e um máximo de 1 desvio-padrão. Imediatamente depois, fizemos uma pergunta semelhante sobre o impacto esperado em outros dois contextos. Esses outros contextos foram escolhidos aleatoriamente entre quatro contextos nos quais estudos acadêmicos estimaram o impacto dos programas de DPI usando experimentos. Os estudos variam em localização e tamanho da amostra. Eles avaliam comparáveis programas de DPI na Colômbia ($n = 1.420$) (Attanasio *et al.*, 2014); na Jamaica ($n = 130$) (Grantham-McGregor *et al.*, 1991; Walker *et al.*, 2005); em Michigan ($n = 123$) (Barnett, 2011); e em vários estados dos Estados Unidos ($n = 4.667$) (Puma *et al.*, 2010). Quando os estudos relevantes foram apresentados ao participante, destacamos o local do estudo e o tamanho da amostra.¹⁶

TABELA 8
Características dos estudos

Atributos	Amostra pequena	Amostra grande
País em desenvolvimento	Jamaica ($n = 130$)	Colômbia ($n = 1.420$)
País rico	Michigan ($n = 123$)	Estados Unidos ($n = 4.667$)

Elaboração dos autores.

Como não é possível observar o verdadeiro efeito da política nos municípios dos participantes, não pudemos incentivar as crenças iniciais dos participantes em termos da sua precisão. No entanto, aleatorizamos incentivos para prever com precisão o efeito nos outros dois contextos – nos quais podemos comparar a previsão do participante com as estimativas dos estudos. Na prática (tabela de resultados não reportada neste capítulo), notamos que o tamanho dos incentivos não tem efeito nenhum sobre as crenças iniciais, DAP ou crenças posteriores, sugerindo que os participantes levaram as questões a sério mesmo na ausência de incentivos; e que fazer melhores previsões com o intuito de obter um retorno maior dentro do experimento não é um fator importante para a interpretação dos nossos resultados.

3.2.3 DAP e atualização de crenças: rodada 1

Depois que os participantes relataram suas crenças iniciais, oferecemos a eles a chance de comprar os resultados (ou seja, obter o tamanho do efeito estimado) de um estudo escolhido aleatoriamente. A moeda de troca experimental que utilizamos para extrair a DAP consistia em bilhetes de loteria, os quais também foram utilizados para incentivar a participação no experimento. Inicialmente, demos a cada

16. Não usamos os rótulos “país em desenvolvimento” ou “país rico”, nem “amostra pequena” ou “amostra grande”. Simplesmente apresentamos a localização e o tamanho da amostra.

participante cem desses bilhetes de loteria.¹⁷ Os participantes poderiam guardar seus bilhetes de loteria para o sorteio ou usar alguns, ou todos, para comprar os resultados das avaliações de impacto.

Seguindo o procedimento de Becker-DeGroot-Marschak (BDM), medimos a DAP máxima do participante [0 a 100] para saber os resultados do estudo em questão.¹⁸ Em seguida, sorteamos um preço aleatório para o estudo. Se o preço estava abaixo da DAP do participante, revelávamos os resultados e deduzíamos o preço sorteado do estoque de bilhetes de loteria do participante. Para lidar com a autosseleção em saber o resultado do estudo com base na própria DAP, e mantendo a compatibilidade de incentivos do procedimento BDM, o preço foi sorteado de uma distribuição com alta massa em zero. Consequentemente, de 80% a 90% (dependendo da conferência) dos participantes receberam o resultado do estudo independentemente de sua DAP. Nas tabelas de resultados de atualização de crenças, apresentamos tanto as estimativas considerando a amostra total quanto as estimativas considerando a subamostra que recebe sem seleção as informações dos estudos (ou seja, participantes que receberam um preço de zero).

Para aqueles participantes que receberam os resultados do estudo, subsequentemente perguntamos novamente suas crenças (posteriores) sobre o efeito esperado da política em seu próprio município, assim como no contexto do estudo que não foi oferecido para compra nessa rodada. Para aqueles que não receberam as informações (ou seja, o preço sorteado foi maior que sua DAP), não perguntamos novamente sobre suas crenças. Como é padrão em experimentos sobre atualização de crenças, presumimos que as crenças não mudam em questão de minutos na ausência de novas informações (Benjamin, 2019; Vivalt e Coville, 2021; Mobius *et al.*, 2011; Eil e Rao, 2011).¹⁹

3.2.4 DAP e atualização de crenças: rodada 2

Na segunda rodada do experimento, apresentamos ao participante um *menu* com os três estudos que não foram oferecidos para compra na rodada 1, novamente destacando a localização de cada estudo e o tamanho da amostra. O participante recebeu um novo orçamento de cem bilhetes de loteria e foi informado de que um dos três estudos seria oferecido de forma aleatória para compra. Solicitamos aos

17. Cada bilhete com uma chance de ganhar uma viagem gratuita para participar de um evento com grande visibilidade política debatendo o Brasil na Universidade de Harvard nos Estados Unidos, onde também participava de um passeio pelo *campus* da universidade.

18. Em um procedimento BDM, um indivíduo relata um lance para um item. O preço do item é logo sorteado aleatoriamente. Se o lance for superior (ou igual) ao preço sorteado, o indivíduo recebe o item e paga o preço sorteado. Se o lance for inferior ao preço sorteado, o indivíduo não recebe o item e não faz pagamento nenhum.

19. Isso elimina a possibilidade de que ser questionado uma segunda vez causaria uma mudança sistemática nas crenças, por exemplo, devido a pensar mais profundamente na questão. Sob nossa suposição, a mudança de crenças é o efeito de ter recebido os resultados do estudo.

participantes que declarassem sua DAP para cada estudo, de forma a comparar a DAP com o preço sorteado caso o estudo em questão fosse escolhido aleatoriamente para venda. Desse modo, obtivemos DAPs incentivo-compatíveis para cada um dos três estudos. Revelamos os resultados do estudo correspondente seguindo o mesmo procedimento anteriormente descrito e, novamente, perguntamos sobre suas crenças (posteriores).

Ter essa segunda rodada nos permite observar uma segunda instância de atualização de crenças por participante, aumentando o poder estatístico da análise. Também nos permite entender como o peso atribuído às informações sobre estudos de avaliação de impacto diminui (ou não) do primeiro para o segundo aprendizado sobre o tema.

3.3 Resultados

Interpretamos os resultados através de uma estrutura simplificada de aprendizagem bayesiana. Suponha que o formulador de política i tenha uma crença inicial $S_i^{pr} \sim \mathcal{N}(\mu_i^{pr}, \Sigma_i^{pr})$, em que μ_i^{pr} é a média das crenças iniciais de i e Σ_i^{pr} , a variância percebida ou incerteza de sua crença inicial sobre o provável efeito da política de DPI, se implementada em seu município. O resultado (ou seja, o tamanho do efeito) do estudo informado pode ser considerado um sinal ruidoso $S_{i,c}^l \sim \mathcal{N}(\mu_l, \Sigma_{i,c}^l)$, extraído de uma distribuição centrada em torno do valor verdadeiro μ_l , mas com variância $\Sigma_{i,c}^l$, em que c indexa características do estudo, como seu tamanho de amostra ou localização. Então, um formulador de política bayesiana que pretende ter crenças precisas (de forma de minimizar o erro quadrático médio) formará uma crença atualizada S_i^{po} :

$$S_i^{po} = (1 - \pi) S_i^{pr} + \pi S_{i,c}^l,$$

com o parâmetro de ponderação $\pi = \Sigma_i^{pr} / (\Sigma_i^{pr} + \Sigma_{i,c}^l)$. Ou seja, a crença posterior de um indivíduo com aprendizado bayesiano será uma combinação convexa de sua crença inicial e do “sinal” (ou seja, o resultado do estudo), com ponderação proporcional à precisão relativa percebida de cada componente. Embora não possamos testar as suposições deste modelo – particularmente que as crenças sigam uma distribuição normal –, uma vez que medimos apenas crenças pontuais, esta estrutura fornece uma referência útil para o processo de atualização de crenças que estudamos com o experimento.

Podemos pensar nos atributos do estudo – localização e tamanho da amostra – como fatores que afetam a precisão percebida ou a informatividade do sinal ruidoso. Se os participantes pensarem que os estudos de amostras maiores são mais informativos ($\Sigma_{i,grande}^l < \Sigma_{i,pequena}^l$), eles colocarão um peso maior no sinal de estudos de amostra grande ao formar sua atualização de crenças. É importante

ressaltar que, se os formuladores de política valorizam ter crenças precisas sobre a efetividade das políticas de DPI, sua DAP pelo resultado de estudos será maior para os resultados daqueles estudos que, *ex-post*, eles ponderarão mais fortemente no seu processo de atualização de crenças.

3.3.1 Crenças iniciais sobre o efeito de DPI

Começamos analisando as crenças iniciais dos formuladores de política sobre a efetividade das políticas de DPI (tabela de resultados não reportada neste capítulo). O formulador de política médio parece sensato, embora um pouco otimista. Em média, o formulador de política da nossa amostra acredita que as políticas de DPI são mais efetivas nos países ricos (tamanho do efeito de 0,45-0,50 desvio-padrão) do que nos países em desenvolvimento (tamanho do efeito de 0,37-0,42 desvio-padrão).

As autoridades municipais acreditam que o efeito em seu próprio município (tamanho do efeito de 0,42 desvio-padrão) está muito próximo da média da crença inicial para os países em desenvolvimento. No entanto, isso esconde uma heterogeneidade substancial nas crenças iniciais: o desvio-padrão das crenças iniciais é de 0,22, o que sugere um desacordo substancial entre os formuladores de política.²⁰ Como apenas medimos crenças pontuais, em vez de crenças probabilísticas completas, não temos uma medida da incerteza nas crenças de cada formulador de política.

3.3.2 Disposição a pagar pelos estudos de DPI

Depois que os formuladores de política relataram suas crenças iniciais, nós solicitamos suas DAPs para saber os resultados das avaliações de impacto de um dos quatro estudos (atribuídos aleatoriamente). Se os formuladores de política valorizam crenças precisas, a DAP deve ser maior quanto mais informativo o sinal for percebido. Estimamos a seguinte equação:

$$DAP_{ijs} = \beta_0 + \beta_1 Desenvolvimento_{ijs} + \beta_2 Grande_{ijs} + \varepsilon_{ijs}, \quad (1)$$

em que DAP_{ijs} é a DAP (em termos de bilhetes do sorteio) do formulador de política i , na rodada $j \in 1, 2$, para comprar o resultado do estudo $s \in$ Michigan, Estados Unidos, Jamaica, Colômbia. $Desenvolvimento_s$ é uma variável *dummy* igual a 1 para estudos na Jamaica ou na Colômbia, e 0 em outros lugares. $Grande_s$ é uma variável *dummy* igual a 1 para os dois estudos de amostra grande (Colômbia com $n = 1.420$ e Estados Unidos com $n = 4.667$), e 0 para caso contrário (Jamaica com $n = 130$ e Michigan com $n = 123$). Os erros-padrão são agrupados no nível individual.

20. É claro que parte do desacordo nas crenças iniciais pode refletir ruído no processo de obtenção de crenças implementado no experimento.

A tabela 9 apresenta os resultados da estimação por MQO da especificação (1). A coluna 1 agrupa as duas rodadas, enquanto as colunas 2 e 3 apresentam separadamente as estimativas para as rodadas 1 e 2, respectivamente. Observamos que os formuladores de política alocam em média 45 bilhetes do sorteio (dos cem bilhetes que recebem em cada rodada) para se informar sobre o efeito de um estudo específico. Embora essa seja uma proporção grande de sua dotação de bilhetes, é difícil interpretar o nível diretamente, pois a moeda experimental é bilhete de loteria, cujo valor subjetivo não é observado.

Para ter uma referência do nível da DAP, oferecemos a uma subamostra de participantes a possibilidade de comprar com os bilhetes do sorteio e utilizando o mesmo procedimento BDM, cartões-presente de uma grande rede de varejo *online* (Lojas Americanas). Com as informações desse exercício, calculamos uma taxa de câmbio entre dinheiro e a moeda experimental. Encontramos que um bilhete adicional foi trocado por aproximadamente US\$ 0,80 em cartões-presente. Esta taxa de câmbio deve ser interpretada com cautela, mas sugere que a DAP média pelos resultados dos estudos é equivalente a US\$ 36 – entre 0,4% e 0,9% do salário mensal de um prefeito.

Há uma heterogeneidade substancial na demanda: o desvio-padrão da DAP é de 32 bilhetes do sorteio. Ainda assim, 99% dos participantes declararam uma DAP estritamente positiva.²¹ A DAP diminui da rodada 1 para a rodada 2: o segundo estudo oferecido a um formulador de política é valorado 11% menos do que o primeiro estudo.

21. Os leitores talvez se perguntem por que os participantes simplesmente não consultam os resultados por sua própria conta. Embora isso possa acontecer até certo ponto, acreditamos que a falta de familiaridade com as fontes de informação de estudos de avaliação de impacto, as barreiras linguísticas e a dificuldade de interpretação da redação acadêmica são fatores que tornam essa estratégia difícil para os participantes do experimento. Nossas estimativas podem ser interpretadas como capturando a DAP por informações simplificadas e convenientemente apresentadas.

TABELA 9
Experimento de crenças: DAP por atributos dos estudos

Variável	(1) DAP	(2) DAP	(3) DAP
Grande	3,8221 (0,7912)	2,3554 (2,3944)	4,4182 (1,0152)
Desenvolvimento	0,3783 (0,7907)	1,5948 (2,3951)	-0,2735 (1,0039)
Observações	2.573	764	1.809
Rodada	1 e 2	1	2
Indivíduos	764	764	604
Média	44,62	48,39	43,03
Desvio-padrão	31,77	33,06	31,09

Elaboração dos autores.

Obs.: Estimativas de MQO. A variável dependente é a DAP, para saber os resultados de estudos de avaliação de impacto de programas de DPI, que são obtidos em duas rodadas diferentes. *Desenvolvimento* é uma *dummy* que é igual a 1 para Jamaica e Colômbia, 0 para caso contrário. *Grande* é uma *dummy* que é igual a 1 para Colômbia e Estados Unidos, 0 para caso contrário. A diferença no número de indivíduos entre as colunas 2 e 3 deve-se em parte a um desenho experimental diferente utilizado na última conferência da CNM, na qual apenas um estudo foi oferecido para compra. Média é a DAP média no lado esquerdo de cada equação. Desvio-padrão é o desvio-padrão da DAP no lado esquerdo de cada equação. Erros-padrão robustos agrupados no nível individual estão entre parênteses. O *p*-valor do teste *grande* (coluna 2) igual a *grande* (coluna 3) é 0,484. O *p*-valor do teste *desenvolvimento* (coluna 2) igual a *desenvolvimento* (coluna 3) é 0,524.

Em seguida, analisamos se a demanda pelos resultados de avaliações de impacto varia com os atributos dos estudos e observamos que os líderes políticos estão dispostos a pagar cerca de 9% a mais por estudos de amostra grande do que por estudos de amostra menor. Os formuladores de política, portanto, parecem valorizar, *ex-ante*, antes da obtenção de informação, a precisão estatística dos resultados de um estudo. Essa relação é mais forte na segunda rodada, quando os estudos são oferecidos lado a lado, mas a estimativa da segunda rodada não é estatisticamente diferente da estimativa da primeira rodada (*p*-valor 0,484). Em contraste, e ao contrário de nossas expectativas iniciais, não encontramos diferenças significativas entre a DAP por resultados de avaliações de impacto da Colômbia e da Jamaica em relação a Michigan ou aos Estados Unidos. Isso sugere que, em média, os formuladores de política municipal no Brasil não consideram os estudos de outros países em desenvolvimento mais informativos – com maior validade externa para eles – do que estudos de países ricos.

Na tabela 10, apresentamos as correlações entre as características dos participantes e dos municípios e suas DAPs. Apenas três características de um total de vinte são significativamente associadas à DAP: i) se o participante é do sexo masculino; ii) se o município onde trabalha o participante já implementou uma política de DPI; e iii) se o participante relatou ter ouvido falar sobre essas políticas, apesar de não as ter implementado em seu município.

Essas últimas duas correlações não são anormais se considerarmos o arcabouço teórico utilizado: os formuladores de política com mais experiência ou conhecimento em determinada política podem ter uma crença inicial mais precisa e, portanto, não valorizar informações adicionais sobre a política. Em vez disso, notamos que são precisamente aqueles formuladores de política que implementam e gastam recursos municipais em programas de DPI que têm maior DAP por informações de estudos sobre o tema. Presumivelmente, isso acontece porque ter crenças precisas sobre esse tipo de programa é mais valioso para esses formuladores. Os prefeitos no segundo mandato e aqueles com maior margem de vitória eleitoral (que presumivelmente enfrentam menor competição eleitoral), por sua vez, não têm menor DAP por informações de estudos de avaliação de impacto.

TABELA 10
Experimento de crenças: DAP por outros determinantes

Variável	(1) DAP	(2) DAP	(3) DAP	(4) DAP
Características dos prefeitos				
Homem	6,74 (3,16)	-	-	6,33 (3,11)
Idade	-0,82 (2,10)	-	-	-0,93 (2,11)
Ensino superior	1,12 (2,16)	-	-	2,45 (2,21)
Segundo mandato	1,16 (2,59)	-	-	1,47 (2,88)
Margem de vitória eleitoral	1,08 (2,13)	-	-	1,45 (2,09)
Partido político de esquerda	0,98 (2,15)	-	-	0,45 (2,22)
Características dos municípios				
População	-	2,64 (2,22)	-	2,05 (2,20)
População com ensino superior	-	-0,52 (2,65)	-	-0,67 (2,68)
Funcionário público com ensino superior	-	2,09 (2,27)	-	0,74 (2,29)
Pobreza	-	-1,41 (5,13)	-	0,32 (5,14)
Gini	-	-0,61 (2,53)	-	-0,88 (2,51)
Grande sul	-	1,93 (4,94)	-	4,93 (5,00)
Renda <i>per capita</i>	-	-5,21 (4,59)	-	-3,80 (4,51)
Crianças na escola (0-3 anos)	-	1,19 (2,34)	-	0,74 (2,34)
Crianças na escola (4-5 anos)	-	2,31 (2,45)	-	2,20 (2,41)

(Continua)

(Continuação)

Variável	(1) DAP	(2) DAP	(3) DAP	(4) DAP
Características questionário etapa introdutória				
Prefeito	-	-	-1,07(2,08)	-0,98(2,16)
Político de carreira	-	-	-0,50(2,34)	-1,40(2,49)
Esquerdista	-	-	0,06(2,50)	0,37(2,54)
Implementou DPI	-	-	11,45(2,39)	11,90(2,47)
Ouviu sobre DPI	-	-	6,84(2,68)	6,89(2,75)
Observações	2.542	2.573	2.573	2.542
Indivíduos	754	764	764	754
Média	44,27	44,62	44,62	44,27

Elaboração dos autores.

Obs.: Estimativas de MQO. A variável dependente é a DAP, para saber os resultados de estudos de avaliação de impacto de programas de DPI, que são obtidos em duas rodadas diferentes. Expressamos todas as variáveis contínuas como indicadores de acima-abaixo da mediana da distribuição dos municípios. As características dos prefeitos incluídas no modelo são: homem (1/0); idade acima ou abaixo da mediana (1/0); ensino superior (1/0); segundo mandato (1/0); margem de vitória eleitoral acima/abaixo da mediana (1/0); e partido político de esquerda (1/0, prefeitos pertencentes a um partido de centro-esquerda de acordo com plataformas políticas históricas). As características dos municípios incluídas no modelo são: população acima/abaixo da mediana (1/0); população com ensino superior acima da mediana abaixo (1/0); funcionário público com ensino superior acima da mediana abaixo (1/0); pobreza acima da mediana (1/0); Gini acima/abaixo da mediana (1/0); Grande sul (1/0, em que 1 são as regiões Sul, Sudeste e Centro-Oeste; e 0 são as regiões Norte e Nordeste); renda *per capita* mensal acima/abaixo da mediana (1/0); crianças na escola (0-3 anos) acima/abaixo da mediana (1/0) da proporção de crianças de 0 a 3 anos de idade no município que frequentam a pré-escola; crianças na escola (4-5 anos) acima/abaixo da mediana (1/0) da proporção de crianças de 4 a 5 anos de idade no município que frequentam a pré-escola. Características do questionário etapa introdutória: prefeito (1/0); político de carreira (1/0); esquerdistas (1/0); implementou DPI (1/0) indica se o participante relatou que o município já implementou um programa de DPI; ouviu sobre DPI (1/0) indica se o participante relatou que já ouviu falar sobre programas de DPI. Média é a DAP média no lado esquerdo de cada equação. Erros-padrão robustos agrupados no nível do indivíduo estão entre parênteses.

3.3.3 Atualização de crenças

Na análise da DAP, observamos que os líderes políticos municipais valorizam os resultados de estudos de avaliação de impacto e pagam mais por estudos que tem uma amostra maior. A seguir, analisamos se os formuladores de política atualizam suas crenças ao serem informados sobre os resultados dos estudos. Observe que, se os formuladores de política compraram os resultados dos estudos apenas para usá-los como motivação para persuadir outras pessoas, por exemplo, eles talvez não atualizem suas próprias crenças ao receber as informações.

Seguindo a estrutura bayesiana, estimamos a seguinte equação:

$$Posterior_{ijs} = \beta_1 Inicial_{ij} + \beta_2 Sinal_{ijs} + \varepsilon_{ij} , \tag{2}$$

em que $Posterior_{ijs}$ é a crença posterior do formulador de política i , na rodada j , sobre o provável efeito em seu próprio município da política de DPI, após receber o resultado do estudo $Sinal_{ijs}$, que tem um tamanho de efeito s . As crenças posteriores após a rodada 1 são utilizadas como crenças iniciais para a rodada 2. Os erros-padrão são agrupados no nível individual.

A tabela 11 apresenta as estimativas de MQO da especificação (2). A coluna 1 agrupa as duas rodadas, enquanto as colunas 2 e 3 apresentam separadamente as estimativas para as rodadas 1 e 2, respectivamente. Consistente com o arcabouço utilizado, $\hat{\beta}_1$ e $\hat{\beta}_2$ são positivos, estatisticamente significantes e somam aproximadamente 1. Os participantes colocam, em média, cerca de dois terços do peso em suas crenças iniciais e um terço no sinal do estudo. Isso indica que eles não simplesmente aceitam ou repetem o sinal do estudo, o que reduz potenciais preocupações sobre efeitos de demanda nas respostas dos participantes.

Os formuladores atribuem um peso semelhante ao sinal do estudo tanto quando atualizam suas crenças em relação ao seu próprio município como quando atualizam suas crenças em relação ao contexto alternativo (coluna 2 *versus* coluna 4). Eles colocam mais peso na sua crença inicial na segunda rodada, quando já incorporaram o achado do primeiro estudo que receberam, do que na primeira rodada. Em outras palavras, o peso atribuído aos resultados de um estudo cai 30% do primeiro para o segundo estudo sobre o qual um formulador de política recebe informações. Conforme descrito anteriormente, pelo desenho do experimento, entre 80% e 90% dos participantes receberam um preço zero e, por conseguinte, receberam as informações sobre os resultados dos estudos independentemente de sua DAP. A coluna 5 restringe a atenção a essas observações e encontra resultados muito semelhantes aos da amostra completa.

TABELA 11
Experimento de crenças: atualização de crenças

Variável	(1) Posterior	(2) Posterior	(3) Posterior	(4) Posterior	(5) Posterior
Inicial	0,6824 (0,0214)	0,5902 (0,0295)	0,7902 (0,0237)	0,6528 (0,0280)	0,6813 (0,0224)
Sinal	0,3230 (0,0194)	0,3749 (0,0261)	0,2607 (0,0234)	0,3622 (0,0293)	0,3209 (0,0203)
Observações	1.240	700	540	543	1.131
Rodada	1 e 2	1	2	1	1 e 2
Crenças sobre	Município	Município	Município	Estudo aleatório	Município
Recebeu estudo gratuitamente	Não	Não	Não	Não	Sim
Indivíduos	755	700	540	543	731

Elaboração dos autores.

Obs.: Estimativas de MQO. A variável dependente é a crença posterior, que é declarada após comprar com sucesso os resultados de um estudo em cada rodada. Inicial é a crença do entrevistado sobre o efeito, logo antes de comprar um estudo. Sinal é o tamanho do efeito do estudo comprado. As crenças posteriores após a rodada 1 são utilizadas como crenças iniciais na rodada 2. Nas linhas abaixo dos coeficientes, crenças sobre específica para que contexto as crenças são solicitadas, seja o próprio município do participante (colunas 1, 2, 3 e 5), seja um dos quatro possíveis contextos dos estudos (coluna 4). Recebeu estudo gratuitamente indica se o participante recebeu as informações dos estudos independentemente de sua DAP. A diferença no número de indivíduos entre as colunas 2, 3, e 4 deve-se em parte a um desenho experimental diferente utilizado na última conferência da CNM, na qual apenas um estudo foi oferecido para compra. Erros-padrão robustos agrupados no nível individual estão entre parênteses.

A análise exploratória de heterogeneidade na atualização das crenças (tabela de resultados não reportada neste capítulo) sugere que prefeitos com formação universitária e esquerdistas colocam menos peso em suas crenças iniciais e mais peso nos resultados dos estudos que seus colegas. Os prefeitos mais velhos fazem o inverso: eles atualizam menos suas crenças quando confrontados com informações de estudos científicos. Embora os prefeitos que implementaram programas de DPI tenham maior DAP por estudos, conforme descrito anteriormente, eles não atualizam mais que outros prefeitos com base nos sinais recebidos. Por fim, os incentivos à reeleição e à competição política não têm uma relação sistemática com a atualização de crenças. Assim como os prefeitos com mandato limitado e aqueles com maiores margens de vitória eleitoral não tiveram uma DAP menor, eles também não atribuem menor peso aos resultados dos estudos. É importante mencionar que essas heterogeneidades são resultados sugestivos e não podem ser interpretadas causalmente.

A fim de testar se os participantes atualizam mais com base em estudos de amostras grandes ou de países em desenvolvimento, estimamos a seguinte equação:

$$\begin{aligned}
 \text{Posterior}_{ijs} = & \beta_1 \text{Inicial}_{ij} + \beta_2 \text{Sinal}_{ij} + \beta_3 \text{Desenvolvimento}_{ijs} \times \\
 & \text{Inicial}_{ij} + \beta_4 \text{Desenvolvimento}_{ijs} \times \text{Sinal}_{ij} + \beta_5 \text{Grande}_{ijs} \times \text{Inicial}_{ij} + \quad (3) \\
 & \beta_6 \text{Grande}_{ijs} \times \text{Sinal}_{ij} + \epsilon_{ij},
 \end{aligned}$$

em que $\text{Desenvolvimento}_{ijs}$ e Grande_{ijs} são definidos como na equação (1). Segundo o nosso modelo, se um indivíduo percebe que um estudo é mais informativo, ele colocará mais peso no sinal desse estudo e, correspondentemente, menos peso em sua crença inicial. Portanto, para testar se os participantes percebem (digamos) estudos de amostras grandes como mais informativos, podemos testar se $\beta_5 < 0$ e $\beta_6 > 0$ ou, em vez disso (um teste mais fraco), se $\beta_6 - \beta_5 > 0$.

A tabela 12 apresenta os resultados de MQO da especificação (3). Novamente, a coluna 1 agrupa as duas rodadas, enquanto as colunas 2 e 3 apresentam separadamente as estimativas de cada rodada. Encontramos evidências consistentes de que os participantes atribuem maior importância aos sinais de estudos de amostras grandes, mas não aos sinais de estudos de países em desenvolvimento. Isso está de acordo com os achados sobre a DAP e confirma que esses formuladores de política consideram os estudos de amostras grandes mais informativos, mas não consideram que os estudos de países ricos e países em desenvolvimento sejam diferentes. O maior peso atribuído aos estudos de amostras grandes fica evidente também na primeira rodada, quando um estudo é apresentado isoladamente. O padrão de resultados se mantém, e de fato é ligeiramente reforçado, quando restringimos a atenção aos casos em que o preço sorteado foi zero (coluna 5).

TABELA 12
Experimento de crenças: atualização de crenças por atributos dos estudos

Variável	(1) Posterior	(2) Posterior	(3) Posterior	(4) Posterior	(5) Posterior
Inicial	0,6388 (0,0368)	0,5600 (0,0531)	0,7509 (0,0471)	0,6685 (0,0543)	0,6420 (0,0384)
Sinal	0,3306 (0,0284)	0,3780 (0,0397)	0,2653 (0,0384)	0,3351 (0,0429)	0,3280 (0,0299)
Inicial * desenvolvimento	-0,0093 (0,0389)	-0,0247 (0,0599)	-0,0106 (0,0477)	-0,0920 (0,0574)	-0,0083 (0,0414)
Sinal * desenvolvimento	0,0091 (0,0349)	0,0082 (0,0515)	0,0189 (0,0472)	0,0682 (0,0578)	0,0039 (0,0367)
Inicial * grande	-0,0535 (0,0480)	-0,0904 (0,0690)	-0,0307 (0,0600)	-0,0563 (0,0714)	-0,0663 (0,0501)
Sinal * grande	0,3233 (0,0712)	0,4068 (0,0963)	0,2413 (0,0942)	0,2744 (0,1176)	0,3510 (0,0745)
Observações	1.240	700	540	543	1.131
Rodada	1 e 2	1	2	1	1 e 2
Crenças sobre	Município	Município	Município	Estudo aleatório	Município
Recebeu estudo gratuitamente	Não	Não	Não	Não	Sim
Indivíduos	755	700	540	543	731
P-valor inicial * desenvolvimento = sinal * desenvolvimento	0,791	0,755	0,742	0,142	0,869
P-valor inicial * grande = sinal * grande	0,001	0,002	0,064	0,069	0,001

Elaboração dos autores.

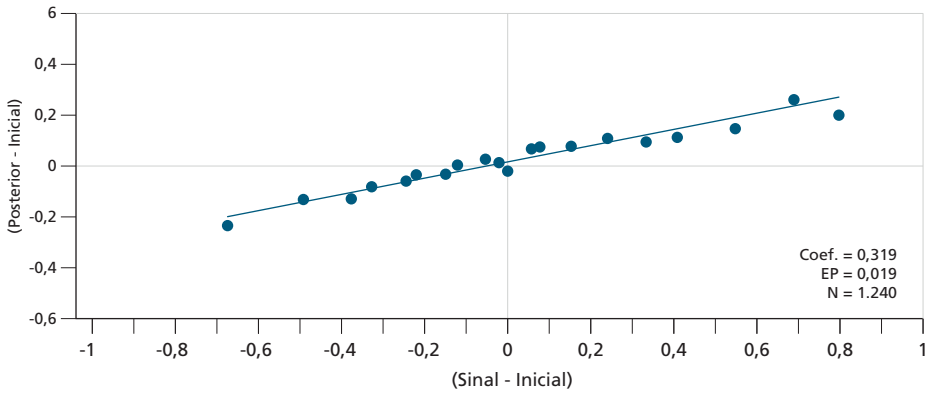
Obs.: Estimativas de MQO. A variável dependente é a crença posterior, que é declarada após comprar com sucesso os resultados de um estudo em cada rodada. Inicial é a crença do entrevistado sobre o efeito, logo antes de comprar um estudo. Sinal é o tamanho do efeito do estudo comprado. As crenças posteriores após a rodada 1 são utilizadas como crenças iniciais na rodada 2. *Desenvolvimento* é uma *dummy* que é igual a 1 para Jamaica e Colômbia, 0 para caso contrário. *Grande* é uma *dummy* que é igual a 1 para Colômbia e Estados Unidos, 0 para caso contrário. Nas linhas abaixo dos coeficientes, crenças sobre específica para que contexto as crenças são solicitadas, seja o próprio município do participante (colunas 1, 2, 3 e 5), seja um dos quatro possíveis contextos dos estudos (coluna 4). Recebeu estudo gratuitamente indica se o participante recebeu as informações dos estudos independentemente de sua DAP. A diferença no número de indivíduos entre as colunas 2, 3, e 4 deve-se em parte a um desenho experimental diferente utilizado na última conferência da CNM, na qual apenas um estudo foi oferecido para compra. Erros-padrão robustos agrupados no nível individual estão entre parênteses.

O gráfico 3, a seguir, mostra a atualização de crenças usando gráficos de dispersão agrupada. O eixo *y* traça o tamanho e a direção da atualização (*Posterior – Inicial*) para um dado choque de informação devido ao sinal (*Sinal – Inicial*) no eixo *x*. O gráfico 3A inclui todas as instâncias de atualização, agrupando estudos e rodadas, e adiciona uma linha de regressão (MQO) ajustada.

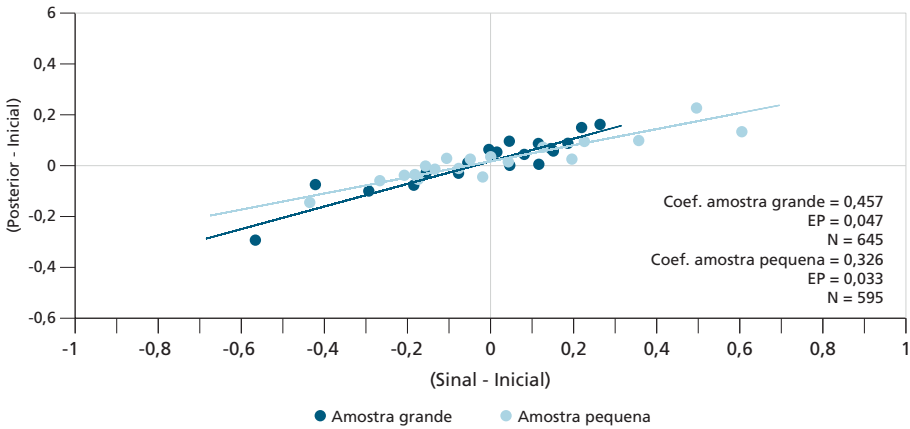
Alguns pontos são importantes de salientar. Primeiro, a relação parece ser linear, conforme o modelo bayesiano e nossa especificação empírica em (2). Segundo, não há evidências de atualização assimétrica (otimista), que apareceria como uma torção na origem, com uma inclinação mais acentuada à direita de zero. Os outros painéis, por sua vez, descrevem a atualização separadamente para estudos de amostra grande e pequena (gráfico 3B) e para estudos de países ricos e em

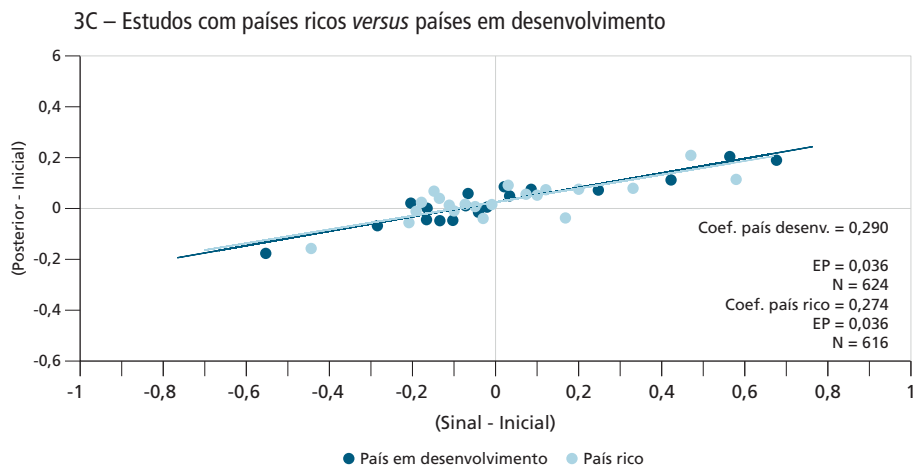
desenvolvimento (gráfico 3C). A atualização mais forte para estudos de amostras grandes é evidente, assim como a atualização semelhante entre estudos de países ricos e em desenvolvimento.

GRÁFICO 3
Experimento de crenças: atualização de crenças
3A – Amostra completa



3B – Estudos com amostras grande *versus* pequena





Elaboração dos autores.

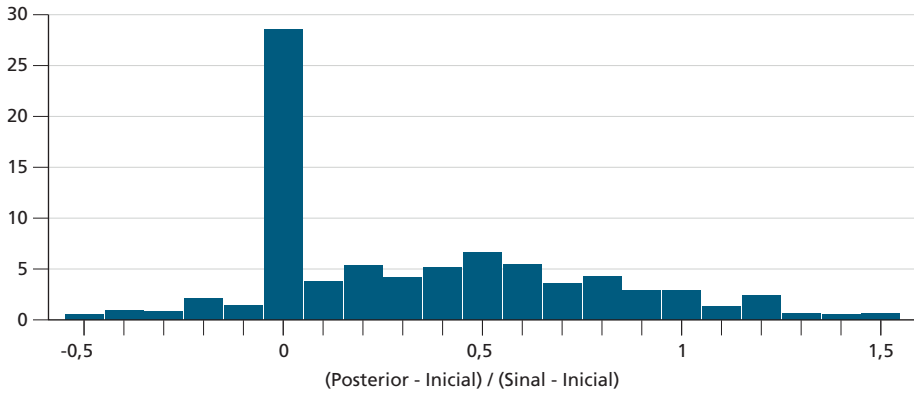
Obs.: 1. Comparação entre a diferença nas percepções dos participantes depois de comprar um estudo (ou seja, crenças posteriores menos crenças iniciais) e a diferença nas percepções dos participantes antes de comprar um estudo (ou seja, sinal menos crenças iniciais), média ponderada da rodada 1 e 2. Inicial é a crença do participante sobre o efeito, logo antes de comprar um estudo. Sinal é o tamanho do efeito do estudo comprado. As crenças posteriores após a rodada 1 são utilizadas como crenças iniciais na rodada 2. O gráfico 3A mostra estatísticas para a amostra completa. Os gráficos 3B e 3C comparam estatísticas entre estudos de amostra grande e pequena (controlando para o atributo de localização do estudo) e entre estudos de países em desenvolvimento e ricos (controlando para o atributo tamanho de amostra do estudo), respectivamente. Estudos de amostra grande (pequena) incluem Colômbia e Estados Unidos (Jamaica e Michigan); enquanto os estudos de países em desenvolvimento (ricos) incluem Colômbia e Jamaica (Michigan e Estados Unidos). A inclinação e os erros-padrão robustos agrupados no nível individual são baseados em uma regressão linear com um termo constante.

2. EP – erro-padrão; N – número de observações.

O gráfico 4 apresenta o histograma das respostas de atualização de crenças. Especificamente, para cada instância de atualização, calculamos $\pi = (\text{Posterior} - \text{Inicial}) / (\text{Sinal} - \text{Inicial})$ e, em seguida, calculamos a média dessas respostas para cada indivíduo. O gráfico 4 revela uma heterogeneidade substancial no peso atribuído aos resultados dos estudos: 28% dos formuladores de política parecem ignorar os resultados dos estudos e não atualizam suas crenças ($\pi = 0$); 43% dos formuladores têm pesos de atualização estritamente entre zero e um; e 15% atualizam na direção errada ($\pi < 0$), enquanto 13% reagem exageradamente ($\pi > 1$). Esta distribuição é bastante semelhante à encontrada em Vivalt e Coville (2021), os quais apresentam aos participantes de um *workshop* de avaliação de impacto do Banco Mundial um estudo hipotético em um exercício de atualização de crenças. Eles também encontram uma parcela substancial de participantes que não atualizam suas crenças ($\pi = 0$), e cerca de 55% exibem ($0 \leq \pi \leq 1$). O peso médio de atualização em nossa amostra, cerca de 0,37 na primeira rodada, também é comparável ao peso médio de 0,5 encontrado por Vivalt e Coville (2021).

GRÁFICO 4

Experimento de crenças: distribuição das atualizações de crenças (Em %)



Elaboração dos autores.

Obs.: Distribuição da proporção da diferença nas percepções do entrevistado depois de comprar um estudo (ou seja, crenças posteriores menos crenças iniciais) e a diferença nas percepções do entrevistado antes de comprar um estudo (ou seja, sinal menos crenças iniciais), média ponderada entre as respostas do indivíduo nas rodadas 1 e 2.

O que explica o fato de que aproximadamente um quarto dos participantes não reagiu aos resultados dos estudos? Uma possibilidade é simplesmente desatenção ou minimização de esforço por parte dos participantes. No entanto, o desenho do experimento garantiu que os participantes registrassem pelo menos brevemente os resultados dos estudos. Ao mesmo tempo, os participantes também foram requeridos a declarar ativamente uma crença posterior durante cada rodada. A interpretação pelas lentes do modelo bayesiano seria, em vez disso, que esses formuladores de política têm crenças iniciais muito confiantes e, portanto, pensam que os estudos não são informativos. Embora isso seja possível, é pelo menos um pouco inconsistente com o fato de que 99% dos participantes declararam uma DAP estritamente positiva.

Um outro fator que poderia explicar a falta de atualização das crenças é um problema de arredondamento na mensuração das crenças no experimento. As crenças só podiam ser relatadas em intervalos de 0,1 desvio-padrão. Assim, atualizações de crenças subjacentes de, por exemplo, 0,46 a 0,54 resultariam em atualização nenhuma, se ambas fossem arredondadas para 0,5. Esses problemas de arredondamento também podem inflar a parcela que parece ter uma reação exagerada, uma vez que a atualização de 0,44 para 0,46 pode ser medida como uma atualização de 0,1. Esta é uma ressalva importante para a interpretação da distribuição de atualização de crenças em nível individual do nosso experimento.

Essa preocupação é provavelmente menor ao medir as respostas médias de muitos participantes, que é onde se foca a maior parte da nossa análise.²²

3.3.4 Ressalvas gerais

Os atributos dos estudos, o tamanho de amostra e o local da avaliação podem estar correlacionados nas mentes dos formuladores de política com variáveis omitidas, como a qualidade das avaliações, com a escala de implementação do programa DPI etc. Para tentar esclarecer isso, implementamos um questionário descritivo com um subconjunto da amostra ($n = 294$). Encontramos que 59% dos formuladores, que, no questionário descritivo, preferiram estudos de amostras grandes, escolheram a precisão estatística como motivo. Curiosamente, uma parcela menor também relatou preferir estudos com amostras grandes, porque é mais provável que avaliassem programas implementados em escala (23%) e pelo governo (15%).

No caso do local do estudo, os resultados do questionário não são tão claros: enquanto os indivíduos que preferiram estudos da Colômbia ou da Jamaica relataram que esses países têm (baixos) padrões de vida e capacidade estatal semelhante à de seus municípios, uma parte substancial dos participantes que relatou preferir os estudos dos Estados Unidos também apontou o (alto) padrão de vida e a capacidade estatal como razões. Uma interpretação é que alguns formuladores de política no Brasil percebem seus municípios como mais próximos dos países em desenvolvimento, enquanto outros os percebem como mais próximos dos países ricos.

Uma fraqueza importante do nosso experimento é que consideramos apenas estudos de três países. O que interpretamos como um “efeito país rico” poderia, em vez disso, ser um “efeito Estados Unidos”: os políticos locais brasileiros podem não valorizar estudos de outros países ricos. Da mesma forma, pode ser que eles atribuam um peso muito maior às evidências do Brasil e as considerem muito mais relevantes do que as evidências da Colômbia ou da Jamaica. A nossa capacidade de explorar essas questões foi limitada devido à falta de estudos comparáveis de mais países, incluindo o Brasil.

Os resultados sobre a atualização das crenças (mas não aqueles da DAP) têm outro potencial problema na interpretação: os dois estudos com amostras grandes, na prática, estimaram efeitos menores. Essa é uma característica que seguem os quatro estudos que usamos, e também é uma característica documentada de forma mais geral na literatura de DPI (Barnett, 2011). Pode acontecer que os participantes simplesmente atualizem mais (em termos proporcionais) em resposta a efeitos pequenos, digamos devido a preocupações sobre um viés de publicação maior em estudos com amostras pequenas, ou porque efeitos grandes parecem implausíveis.

22. Os resultados não mudam se eliminarmos os participantes que nunca atualizam ($\pi = 0$), ou que têm $\pi \leq 0$ ou $\pi \geq 1$ (tabela de resultados não reportada neste capítulo).

Durante a implementação do experimento, tivemos algumas variações não planejadas que podem lançar luz sobre esta preocupação: em seis das quatorze conferências nas quais o experimento foi implementado, relatamos diferentes tamanhos de efeitos para certos estudos. Especificamente, apenas para os estudos de amostras pequenas, relatamos os efeitos estimados em um horizonte de tempo muito mais longo (sem sinalizar essa discrepância), o que resultou em um sinal informado menor. Aproveitamos essa variação para testar se o peso nos sinais de estudos de amostra grande é menos pronunciado nessas conferências. Consistente com nossa interpretação inicial, o peso colocado no tamanho da amostra não varia significativamente entre essas conferências (tabela de resultados não reportada neste capítulo).

3.3.5 DAP por informações de implementação

O acesso a resultados de estudos científicos leva à adoção de políticas mais efetivas? No final do experimento de crenças, os participantes tiveram a oportunidade de adquirir informações práticas sobre como implementar políticas de DPI, usando um novo orçamento de bilhetes de loteria. Interpretamos a DAP por essas informações práticas como uma *proxy* de preferência revelada sobre o interesse na implementação da política. Uma vez que variamos experimentalmente os estudos e, portanto, os atributos – tamanho do efeito, local da avaliação e tamanho da amostra – e notamos que afetam as crenças posteriores, podemos usar esses atributos como instrumentos para as crenças posteriores dos participantes.

A tabela 13 mostra os resultados, nos quais observamos que crenças mais positivas sobre programas de DPI – formadas por meio da recepção dos sinais dos estudos de avaliação de impacto – aumentam causalmente a DAP por informações de implementação. Isso fornece evidências experimentais claras sobre os efeitos que as informações de estudos de avaliações de impacto podem ter, via mudança de crenças, na demanda por implementação de políticas.

TABELA 13
Experimento de crenças: DAP por reporte de implementação

Variável	MQO	MQ2E
	(1) Reporte de implementação	(2) Reporte de implementação
Crença posterior final	16,4909 (5,2677)	41,1943 (21,0608)
Observações	737	737
Instrumentalizou posterior final	Não	Sim
Indivíduos	737	737
Média	59,72	59,72
Desvio-padrão	33,69	33,69

Elaboração dos autores.

Obs.: Estimativas de MQO (coluna 1) e de MQ2E (coluna 2). A variável dependente é a disposição de pagar por informações práticas sobre como implementar políticas de DPI. Crença posterior final é a última crença atualizada, ou seja, após a compra de um ou dois resultados. Na coluna 2, esta última variável é instrumentalizada usando o sinal recebido ou a média dos sinais recebidos, caso o participante tenha comprado dois estudos. Média é a DAP média no lado esquerdo de cada equação. Desvio-padrão é o desvio-padrão da DAP no lado esquerdo de cada equação. Erros-padrão robustos agrupados no nível do indivíduo estão entre parênteses.

3.4 Experimento de crenças: discussão

Este experimento tem dois resultados principais. Em primeiro lugar, os líderes municipais do Brasil valorizam o aprendizado de resultados sobre a efetividade de políticas públicas. Em segundo lugar, eles também mudam suas crenças quando confrontados com evidências de estudos de avaliação de impacto, pois colocam um peso substancial nas novas informações – nesse caso mais peso em estudos de amostras maiores, mas não em estudos de países em desenvolvimento.

O experimento tem alguns pontos fracos notórios. Comparamos os efeitos de um número limitado de estudos de apenas três países. A medida da DAP é um tanto artificial e sai do orçamento privado do formulador de política, em vez de sair do orçamento municipal, o que provavelmente seria mais relevante ao ter, por exemplo, diferentes usos alternativos. Estimamos efeitos sobre as crenças considerando um período muito curto de tempo, pelo que não podemos estabelecer a persistência dos efeitos no tempo. O experimento de adoção de políticas descrito anteriormente, em contraste, fornece evidências de mudança nas crenças em um período de tempo maior (15-24 meses) e descreve efeitos na política municipal real.

4 CONCLUSÃO

Políticas públicas são fundamentais para o desenvolvimento econômico. Qual o papel que a pesquisa científica sobre a efetividade de políticas públicas pode desempenhar para estimular a disseminação de políticas efetivas e o abandono de políticas não efetivas? Uma possibilidade é que a falta de (acesso a) informações

sobre evidências de avaliações de impacto não seja uma restrição à escolha de políticas, por exemplo, porque: as evidências são contestadas e não existe uma visão única de como se deve resolver um problema; as soluções não são viáveis ou válidas, dado o contexto local ou temporal; o processo político exige que os líderes tomem decisões rápidas, sem conceder o tempo necessário para processar todas as informações disponíveis; os líderes têm interesse próprio, e as pressões competitivas eleitorais são muito fracas para motivar o esforço necessário para mudar a política; ou, simplesmente, os líderes possuem poder de decisão limitado sobre as políticas públicas em uso. Alternativamente, a existência de fricções de informação pode restringir o acesso dos líderes políticos às pesquisas existentes.

Neste capítulo, investigamos experimentalmente como informar os resultados de estudos de avaliações de impacto afeta as crenças e práticas de política pública dos formuladores de políticas municipais no Brasil. Mostramos que fornecer aos prefeitos resultados de estudos que documentam o impacto positivo de uma política barata e fácil de implementar aumenta a probabilidade de que seus municípios implementem a política. Também mostramos que os líderes locais valorizam o acesso às avaliações de impacto e atualizam suas crenças quando informados sobre os resultados de estudos científicos. Tornar as informações de estudos de avaliações de impacto facilmente disponíveis e entendíveis para os formuladores de políticas públicas parece influenciar as políticas adotadas.

Isso sugere que fricções de informação podem desempenhar um papel importante na adoção de políticas públicas efetivas no nível municipal. Surpreendente que tais falhas de informação persistam. Afinal, mesmo que os próprios líderes políticos não leiam revistas acadêmicas, as barreiras de informação deveriam gerar incentivos para que os atores interessados em melhorar o bem-estar social conectem as pesquisas acadêmicas com os formuladores de políticas de forma a ultrapassar essas barreiras. Porém, é custo-efetivo tentar fazer essa conexão?

Com relação à análise de custo-benefício, exploremos como exemplo o fornecimento de informações de evidências empíricas sobre a efetividade de cartas-lembrado do experimento de adoção de política. De acordo com os cálculos e suposições explicitados a seguir, por cada real gasto em produzir e fornecer as evidências dos estudos se teriam gerado R\$ 27 a mais no orçamento dos municípios.

Os custos para produzir as informações foram de aproximadamente R\$ 4,7 mil em preços de 2016.²³ Junte-se a isso os custos para apresentar as informações,

23. Estes correspondem aos custos de cem horas de um pesquisador com mestrado – considerando o salário-hora pago por organizações internacionais de desenvolvimento – e de desenho do documento-resumo por parte de um *designer* profissional. As horas do pesquisador se distribuíram aproximadamente em sessenta horas para revisão de literatura, vinte horas para escritura de documento resumo e vinte horas para elaboração de dispositivos para a apresentação.

que foram de aproximadamente R\$ 29 mil em preços de 2016.²⁴ No total, 334 prefeitos participaram das nossas sessões de informação. Consequentemente, o custo total de produzir e fornecer as informações foi de aproximadamente R\$ 100 por prefeito participante. Claro que existem custos adicionais fora do nosso orçamento, por exemplo, a passagem aérea dos prefeitos a Brasília. Entretanto, as oportunidades de eventos que já reúnem gestores e políticos no Brasil são inúmeras. Apenas o nosso parceiro, a CNM, organiza vários eventos como esse anualmente, além de outras associações de setores específicos como o Conselho Nacional de Secretarias Municipais de Saúde (Conasems), na saúde, e a União Nacional dos Dirigentes Municipais de Educação (Undime) e o Conselho Nacional de Secretários de Educação (Consed), na educação. É possível, portanto, continuar provendo informação apenas com o custo marginal que explicitamos aqui.

Para estimar os benefícios das sessões de informação, usamos primeiramente os resultados obtidos no estudo: as sessões de informação aumentam em 10,3 p.p. a adoção de cartas-lembrete em comparação a uma base de uso de 31,7 p.p. sem o fornecimento de evidências sobre a efetividade das cartas-lembrete (tabela 3, coluna 1). Além disso, no cálculo, tomamos como base o município médio da amostra do experimento de adoção de política. A média das receitas tributárias totais entre 2010 e 2015 do município médio do experimento foi de aproximadamente R\$ 4.056.344 em preços de 2016. Combinamos a isso as seguintes suposições: i) a política de cartas-lembrete aumenta a conformidade tributária em 12% (principal estimativa compartilhada com os prefeitos durante as sessões de informação) e esse aumento na conformidade se traduz em um aumento de 12% na arrecadação tributária; e ii) todos os municípios que adotam a política aumentam sua arrecadação tributária em 12%.²⁵

Sob este cenário, por cada real gasto em produzir e fornecer as evidências dos estudos, o conjunto dos municípios do grupo de tratamento teria gerado R\$ 1.274 a mais em arrecadação tributária. Se adicionarmos aos custos de produção e fornecimento da informação os custos variáveis de implementação da política de cartas-lembrete (por exemplo, impressão das cartas, envelopes, envio das cartas),

24. Nessa categoria incluíram, entre outros, a contratação de um instrutor brasileiro com experiência internacional por dez dias – considerando o salário-hora pago por organizações internacionais de desenvolvimento –, o aluguel de uma sala privada dentro da conferência para implementar as sessões de informação, as cópias do documento-resumo utilizando uma impressão de alta qualidade e gastos de viagem do instrutor (passagem aérea, hotel etc.).

25. Note que um aumento de conformidade tributária não implica necessariamente um aumento de arrecadação. É possível que apenas a data em que os tributos são pagos se modifique, aumentando a conformidade sem aumentar arrecadação.

ter-se-iam gerado R\$ 27 a mais por cada real gasto.²⁶ Se os prefeitos decidissem focar as cartas-lembrète unicamente em aumentar a conformidade tributária do IPTU, por exemplo, teriam sido gerados R\$ 296 a mais por cada real gasto – sem custo de implementação – e R\$ 6 a mais com custo de implementação.

Estas estimativas sugerem que aproximar o conhecimento produzido nos centros de pesquisa e as demandas dos gestores públicos municipais pode ser altamente custo-efetivo. Eventos que reúnem lideranças e técnicos municipais, assim como parcerias entre municípios e centros de ensino e pesquisa, são espaços naturais de compartilhamento e transmissão de conhecimentos que (potencialmente) podem ser utilizados para aprimorar a formulação de políticas na administração pública municipal. De fato, a CNM já desenvolve iniciativas tentando fazer essa junção – iniciativas como o Prêmio MuniCiência do projeto UniverCIDADES, que conecta universidades com municípios em seu redor.²⁷

Acreditamos que há, entretanto, um potencial enorme que atualmente não é explorado indo além da influência localizada da atuação de universidades em seus arredores geográficos. Existe um volume de conhecimento substancial produzido por universidades internacionais e nacionais sobre a efetividade de políticas e práticas específicas que atualmente não chega nos tomadores de decisão municipal. A consolidação desse conhecimento e fornecimento direto para gestores não é papel usualmente feito por pesquisadores, e mais, não necessitaria da atuação de pesquisadores – isto é, que atuam criando conhecimento –, mas apenas de pessoas capazes de “consumir” e compreender pesquisa.

Algumas organizações já empreendem o papel de resumir e traduzir tais conhecimentos, como o The Abdul Latif Jameel Poverty Action Lab (JPAL) e o Innovations for Poverty Action (IPA). Entretanto, em sua maioria, atuam (e foram fundadas) para servir pesquisadores a constituir parcerias com gestores com a finalidade de gerar mais conhecimento. Entidades que têm o foco central em política pública podem fazer essa consolidação e beneficiar muito seus clientes e membros, pois terão como objetivo a política pública em si, e não a produção de conhecimento. Iniciativas recentes como A Ponte vão exatamente nessa direção e são muito bem-vindas.²⁸

26. Para simplificar a estimativa dos custos de implementação da política de cartas-lembrète, consideramos um custo fixo igual a zero – uma suposição bastante plausível, tendo em conta o tipo de política. Para estimar os custos variáveis, pressupomos que as prefeituras enviam uma carta-lembrète a todos os domicílios do município. Segundo o IBGE, em cada domicílio brasileiro, moram, em média, aproximadamente três pessoas. Como o município médio da amostra do experimento tem uma população de 21 mil habitantes, pressupomos que as prefeituras enviam 7 mil cartas-lembrète. Levando em conta, em preços de 2016, um custo de R\$ 0,13 por carta impressa, de R\$ 0,85 por envelope tipo comercial e de R\$ 1,5 por carta enviada, o custo total de enviar uma carta-lembrète a todos os domicílios de um município é de aproximadamente R\$ 17.360.

27. Disponível em: <<https://bit.ly/3z9qxkv>>.

28. Ver o *site* da iniciativa A Ponte: <<https://bit.ly/34gSzz8>>.

Podem existir diferentes barreiras à adoção de políticas públicas custo-efetivas em municípios brasileiros, desde, por exemplo, a falta de capacidade técnica até a falta de incentivos políticos. Os resultados discutidos neste capítulo que focaram em duas políticas, uma sobre arrecadação tributária e outra sobre desenvolvimento da primeira infância, demonstram que (a falta de) acesso a evidências científicas é uma barreira relevante na elaboração e implementação de políticas públicas no nível local. Ainda mais, mostramos que a comunicação direta com gestores municipais foi um instrumento efetivo para tornar a política pública com base em evidência uma realidade em municípios brasileiros.

REFERÊNCIAS

ATTANASIO, O. *et al.* Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. **BMJ**, n. 349, 2014.

BARNETT, W. S. Effectiveness of early educational intervention. **Science**, v. 333, n. 6045, p. 975-978, Aug. 2011.

BENJAMIN, D. J. Errors in probabilistic reasoning and judgment biases. *In: Handbook of behavioral economics: applications and foundations 1*. Elsevier, 2019. v. 2. p. 69-186.

CASTRO, L.; SCARTASCINI, C. Tax compliance and enforcement in the Pampas evidence from a field experiment. **Journal of Economic Behavior and Organization**, v. 116, p. 65-82, 2015.

COLEMAN, S. **The Minnesota income tax compliance experiment: state tax results**. Minnesota Department of Revenue, 1996.

DE CESARE, C.; SMOLKA, M. **Diagnóstico sobre o IPTU**. Lincoln Institute of Land Policy, 2004.

DEL CARPIO, L. **Are the neighbors cheating?** Evidence from a social norm experiment on property taxes in Peru. Princeton: Princeton University, 2013. (Working Paper, n. 8544).

EIL, D.; RAO, J. M. The good news-bad news effect: asymmetric processing of objective information about yourself. **American Economic Journal: Microeconomics**, v. 3, n. 2, p. 114-138, 2011.

FELLNER, G.; SAUSGRUBER, R.; TRAXLER, C. Testing enforcement strategies in the field: Threat, moral appeal and social information. **Journal of the European Economic Association**, v. 11, n. 3, p. 634-660, 2013.

GRANTHAM-MCGREGOR, S. M. *et al.* Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican study. **Lancet**, v. 338, n. 8758, p. 1-5, 1991.

HALLSWORTH, M. *et al.* The behavioralist as tax collector: using natural field experiments to enhance tax compliance. **Journal of Public Economics**, v. 148, p. 14-31, 2017.

HASSELDINE, J. *et al.* Persuasive communications: tax compliance enforcement strategies for sole proprietors. **Contemporary Accounting Research**, v. 24, n. 1, p. 171-194, 2007.

HJORT, J. *et al.* How research affects policy: experimental evidence from 2,150 Brazilian municipalities. **American Economic Review**, v. 111, n. 5, May 2021.

MOBIUS, M. M. *et al.* **Managing self-confidence**: theory and experimental evidence. National Bureau of Economic Research, 2011. (Working Paper, n. 17014).

PUMA, M. *et al.* **Head start impact study**: final report. Washington, United States: Office of Planning, Research and Evaluation; U.S. Department of Health and Human Services; Administration for Children & Families, 2010.

VIVALT, E.; COVILLE, A. **How do policymakers update their beliefs?** 2021. Disponível em: <<https://bit.ly/3GuKN2U>>.

WALKER, S. P. *et al.* Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. **Lancet**, v. 366, n. 9499, p. 1804-1807, 2005.

