

## REPLICAÇÃO, REPLICACÃO<sup>1 \*</sup>

Gary King<sup>\*\* 2</sup>

A ciência política é uma empreitada coletiva. A comunidade dos cientistas políticos empiricamente orientados precisa ter acesso aos conjuntos de dados necessários para reproduzir os estudos existentes de modo a entender, avaliar e, especialmente, se desenvolver a partir destes trabalhos. Infelizmente, as regras que nós seguimos atualmente não encorajam, ou em alguns casos nem permitem, este objetivo. A seguir estão sugestões que devem facilitar a replicação\*\*\* e que são facilmente implementáveis por professores, estudantes, pós-graduandos, programas de pós-graduação, autores, pareceristas, agências de financiamento e editores de livros e revistas.

### Problemas na Ciência Política empírica

Como praticamente todo bom livro de metodologia explica, *a única maneira de compreender e avaliar completamente uma análise empírica é conhecer o processo exato pelo qual os dados foram gerados e a análise produzida*. Sem a documentação adequada, os pesquisadores muitas vezes têm dificuldade para replicar os seus próprios resultados meses depois. Uma vez que informações suficientes geralmente estão em falta na ciência política, tentar reproduzir os resultados dos outros, mesmo com a sua própria ajuda, é muitas vezes impossível.

Tanto para análises quantitativas como qualitativas precisamos responder a questões como estas: como os entrevistados foram selecionados? Quem realizou as entrevistas? Qual foi a ordem das questões? Como você decidiu quais informantes entrevistar ou quais cidades visitar? Quanto tempo você passou em cada comunidade? Você falou com as pessoas em sua língua ou por meio de um tradutor? De qual versão do banco de dados do *Interuniversity Consortium for Political and Social Research* (ICPSR) você extraiu as informações? Quão experientes eram os responsáveis pela codificação? Quão frequentemente estes indivíduo

<sup>1</sup> DOI deste artigo: 10.5380/recp.v6i2.44206.

\* Artigo originalmente publicado sob o título “Replication, Replication” na *PS: Political Science & Politics* 28: p. 444–452 (1995). O comitê editorial da *Revista Eletrônica de Ciência Política* agradece ao autor, Gary King, pela autorização para publicação desta tradução. Tradução de Márcio Carlomagno. Revisão técnica de Adriano Codato.

\*\* Gary King é professor do Departamento de Governo da Universidade de Harvard. O autor pode ser contatado em Littauer Center, North Yard, Cambridge, Massachusetts 02138; e-mail: king@harvard.edu.

<sup>2</sup> Este artigo foi beneficiado imensamente por inúmeras conversas que tive com muitos grupos e indivíduos. Por muitos comentários úteis a versões anteriores deste artigo, sou especialmente grato a Jim Alt, Neal Beck, Robert X. Browning, John DiJulio, John Green, Matthew Holden, Gary Klass, David Laitin, Malcolm Litchfield, Ken Meier, Jonathan Nagler, Bob Putnam, Richard Rockwell, Phil Schrodt, e Sid Verba. Agradeço também a National Science Foundation pelas verbas SBR-9321212 e SBR-9223637, e a John Simon Guggenheim Memorial Foundation por uma bolsa.

\*\*\* Nota do Tradutor. Optamos pela manutenção do termo “replicação”, no sentido de repetição/reprodução, mesmo que tal significado ainda não esteja dicionarizado em língua portuguesa, por este conceito já estar consagrado na área da ciência política.

concordaram? Exatamente quais códigos foram originalmente gerados e quais foram as recodificações realizadas? Qual foi exatamente a medida de desemprego utilizada? Quais foram as regras exatas usadas para conduzir a análise de conteúdo? Quando a série temporal começa e acaba? Quais países foram inclusos em seu estudo e como eles foram escolhidos? Quais procedimentos estatísticos foram usados? Qual método de otimização numérica você escolheu? Qual programa de computador foi usado? Como você preencheu ou deletou dados ausentes?

Producir uma lista abrangente de tais questões para cada autor responder, ou decidir *ex ante* quais perguntas poderão ser importantes, é praticamente impossível. Por este motivo, os pesquisadores de orientação quantitativa na maioria das disciplinas têm quase uniformemente adotado o mesmo método de verificar se existe informação suficiente em um trabalho publicado. O *padrão de replicação* entende que existem informações suficientes com as quais seja possível compreender, avaliar e construir a partir de um trabalho anterior quando um terceiro indivíduo poderia reproduzir os resultados sem que o autor precisasse fornecer qualquer informação. O padrão de replicação na verdade não exige que ninguém reproduza os resultados de um artigo ou livro. Ele só requer que informação suficiente seja fornecida no artigo ou livro ou de alguma outra forma acessível ao público, de modo que os resultados possam, a princípio, ser reproduzidos. Uma vez que muitos acreditam que os padrões de pesquisa devem ser aplicados igualmente a análises quantitativas e qualitativas (King, Keohane e Verba, 1994), o padrão de replicação também é apropriado para a pesquisa qualitativa, embora a riqueza e a complexidade destes dados muitas vezes torne isso mais difícil.<sup>3</sup>

O processo de redução dos fenômenos do mundo real para um trabalho científico publicado envolve duas fases: a representação do mundo real por meio de dados descritivos, tanto quantitativos quanto qualitativos, e a análise desses dados. Ambas as fases são componentes importantes do padrão de replicação. Os futuros pesquisadores, apenas com a sua publicação e outras informações que você forneceu, devem ser capazes de, a partir do mundo real, chegar às mesmas conclusões específicas. Em muitos tipos de pesquisa isso não

---

<sup>3</sup> Em alguns casos, o padrão de replicação refere-se à execução das mesmas análises sobre os mesmos dados para chegar ao mesmo resultado, o que provavelmente deve ser chamado de "duplicação" ou talvez "confirmação". Para outros artigos, tal padrão, na verdade, envolve o que é mais usualmente chamado de "replicação" - ir de volta ao mundo a partir do qual os dados vieram e administrar as mesmas medidas, como instrumentos de pesquisa. Uma vez que este envolve diferentes resultados numéricos, devido a uma mudança no tempo, lugar ou sujeitos, não seria de se esperar a duplicação dos resultados publicados exatamente; no entanto, este procedimento permite o benefício científico de verificar se as conclusões substanciais são características sistemáticas do mundo ou características idiossincráticas de medição do último autor. Neste artigo, sigo a prática corrente comum nas ciências sociais de se referir a todos esses procedimentos como "replicação".

é possível, mas deve sempre ser tentado. Em princípio, o padrão de replicação pode às vezes ser atendido mesmo sem tornar públicos os dados utilizados na análise, desde que a própria descrição de ambas as fases da análise seja suficientemente detalhada. No entanto, fornecer este nível de detalhe sem os dados é difícil, senão impossível, e bem menos útil para futuros pesquisadores. Além disso, pode não ser possível repetir a fase de coleta de dados, na medida em que o mundo pode ter mudado no momento em que um futuro pesquisador realize o esforço de duplicação.

Um excelente exemplo de um recente estudo sobre a adesão ao padrão de replicação é o de Dewald, Thursby e Anderson (1986). Um dos autores foi o editor do *Journal of Money, Credit and Banking*. Eles receberam uma verba da National Science Foundation para replicar os resultados de todos os artigos aceitos no ano anterior. Seu trabalho é um relatório revelador (e desconcertante) de suas extensas, mas em grande parte fracassadas, tentativas de replicar cada um dos artigos. Seus resultados (p. 587-588) "sugerem que erros involuntários em artigos empíricos publicados são um lugar-comum, em vez de uma ocorrência rara." Mesmo quando eles não encontraram erros, a replicação foi muitas vezes impossível, mesmo com a ajuda do autor original - e a ajuda dos autores muitas vezes não foi fornecida. Mais importante, quando os editores começaram a exigir dos autores que atendessem o padrão de replicação, eles (p. 589) "descobriram que o próprio processo de autores compilarem seus programas e dados para a submissão lhes revelou ambiguidades, erros e descuidos que de outra forma não seriam detectados." Uma vez que cientistas políticos coletam dados muito mais *originais*, em vez de seguir a prática de economistas de confiar principalmente em dados existentes de fontes governamentais, os benefícios de uma política de replicação em nossa disciplina devem ser ainda mais substanciais do que indicado nas conclusões de Dewald, Thursby e Anderson.<sup>4</sup>

Como este exemplo bastante impressionante demonstra, a falha generalizada em aderir ao padrão de replicação coloca sérios problemas para qualquer disciplina. Na sua mais fundamental, se a base empírica de um artigo ou livro não pode ser reproduzida, de que serve à disciplina as suas conclusões? A qual propósito um artigo como este serve? No mínimo, alguma proteção deve ser realizada para impedir os pesquisadores de desperdiçar seu tempo lendo estas obras. Na pior das hipóteses, muito mais tempo pode ser desperdiçado em esforços infrutíferos para expandir, ampliar e construir sobre um corpo de trabalho que não tem fundamento empírico.

---

<sup>4</sup> Para outros trabalhos sobre replicação, a partir das perspectivas de outras ciências sociais, considerações éticas, as vantagens para a ciência, incentivos aos investigadores e outras preocupações, ver Sieber 1991, Ceci e Walker, 1983; Neuliep, 1991; Fienberg et al, 1985; e Feigenbaum e Levy, 1993.

De modo mais geral, o padrão de replicação permite que pesquisadores compreendam e avaliem melhor a pesquisa existente, e selecionem mais discriminadamente entre este corpo de trabalho no desenvolvimento de suas próprias agendas de pesquisa. Sem informações completas sobre de onde os dados vêm e como nós medimos o mundo real e abstraímos a partir disto, nós não podemos compreender verdadeiramente um conjunto de resultados empíricos.<sup>5</sup> Avaliação, da mesma forma, requer, pelo menos, tanta informação quanto. Assim, tanto pareceristas quanto editores de revistas e livros devem ser providos com informações suficientes para replicar o trabalho que lhes foi submetido para publicação. Talvez o mais importante, o padrão de replicação é extremamente importante para o desenvolvimento da disciplina. O método mais comum e cientificamente produtivo de construção em pesquisa existente é replicar um achado existente - seguindo o caminho exato tomado por um pesquisador anterior, e em seguida, melhorar os dados ou metodologia de uma forma ou outra. Este procedimento assegura que o segundo pesquisador receberá todos os benefícios do trabalho duro do primeiro pesquisador. Afinal, é por isso que os acadêmicos referem-se a artigos e livros como "*contribuições acadêmicas*", e essas contribuições são reconhecidas com citações, agradecimentos, promoções e aumentos salariais. Tais contribuições são consideravelmente mais valiosas quando o custo de construção a partir dela é tão pequeno quanto possível.

Reproduzir e, em seguida, ampliar pesquisa existente de alta qualidade é também uma ferramenta pedagógica extremamente útil, embora os estudantes de ciência política apenas raramente tenham sido capazes de explorar isto, dada a limitada adesão da disciplina ao padrão de replicação. Além disso, para além destas razões altruístas para apoiar o padrão de replicação, há uma motivação adicional, mais auto-interessada: um artigo que não pode ser replicado geralmente é lido com menos frequência, citado com menos frequência e pesquisado menos exaustivamente por outros acadêmicos. Poucos eventos na vida acadêmica são mais frustrantes do que investir enormes quantidades de tempo, esforço e orgulho em um artigo ou livro, apenas para tê-lo ignorado pela profissão, não seguido por outros pesquisadores, não utilizado para construir outras pesquisas bem-sucedidas ou não explorado em outros contextos. Além disso, ser ignorado é muito prejudicial para a carreira, mas ser aplaudido, citado favoravelmente, criticado ou até mesmo atacado são todas evidências igualmente fortes de que você está sendo levado a sério por suas contribuições para o debate acadêmico (veja Feigenbaum e Levy, 1993, citando Diamond 1988, e Leimer

---

<sup>5</sup> Vale a pena mencionar que eu duvido que a fraude seja um grande problema na pesquisa em ciência política. Provavelmente existe até certo ponto, como acontece em qualquer outra disciplina e área do esforço humano, mas não vejo nenhuma evidência de que seja qualquer coisa que não extremamente rara.

e Lesnoy, 1982). Infelizmente, um estudo recente indica que o número modal de citações de artigos em ciência política é *zero*: 90,1% dos nossos artigos nunca são citados! (Hamilton 1991; Pendlebury, 1994) Uma fração ainda menor de artigos estimula a investigação ativa por outros pesquisadores.

Este problema limita muito o nosso conhecimento coletivo de governo e política. A academia é um empreendimento social que normalmente é mais bem sucedido quando os pesquisadores individuais competem e colaboram para contribuir em busca de objetivos comuns. Em contraste, quando trabalhamos isoladamente em problemas não relacionados, ignorando o trabalho que veio antes, perdemos os benefícios de avaliar o trabalho uns dos outros, analisando o mesmo problema a partir de diferentes perspectivas, aperfeiçoando técnicas e métodos, e, mais importante, construindo com base no trabalho existente em vez de repetidamente reinventar a roda.

### Soluções propostas

Soluções para muitos dos problemas existentes na ciência política empírica são melhor implementadas por autores individuais. Contudo, a experiência em muitas disciplinas mostrou que algumas regras formais também são necessárias. Acadêmicos, administradores, pareceristas e editores podem desempenhar um papel importante em encorajar ou exigir a adesão ao padrão de replicação.

### Autores

Se autores individuais desejam aumentar a probabilidade de que seu trabalho será lido, compreendido e levado a sério em pesquisas futuras, seguir o padrão de replicação é um passo muito importante (é também um meio eficaz de assegurar que os pesquisadores serão capazes de dar seguimento a seu próprio trabalho após os pormenores metodológicos tenham se apagado da memória).

Na prática, seguir o padrão de replicação pode envolver colocar mais informações em artigos, livros ou dissertações sobre o processo exato pelo qual a informação foi extraída ou os dados recolhidos, codificados, analisados e relatados. Infelizmente, revistas e livros em geral não irão fornecer espaço suficiente para fazer isso corretamente. Além disso, grande parte do material necessário é melhor comunicado eletronicamente, em vez de em papel. Felizmente, dois dos melhores arquivos digitais da disciplina, descritos mais detalhadamente a seguir, podem facilmente ser usados para satisfazer o padrão de replicação: a coleção do

*Public Affairs Video Archive* da Universidade Purdue e o *Inter-University Consortium for Political and Social Research* da Universidade de Michigan.

O primeiro passo na implementação do padrão de replicação é criar um conjunto de dados de replicação. Estes conjuntos de dados incluem todas as informações necessárias para reproduzir os resultados empíricos. Para os pesquisadores quantitativos, estas informações podem incluir dados originais, programas de computador especializados, conjuntos de recodificação de softwares, extratos de dados públicos disponíveis (ou instruções muito claras sobre como obter exatamente as mesmas que você usou) e uma nota explicativa (geralmente na forma de um arquivo "leia-me") que descreva o que está incluído e explique como reproduzir os resultados numéricos do artigo. Não é necessário fornecer todas as informações disponíveis, apenas o subconjunto de variáveis e observações do banco de dados usado para produzir os resultados publicados.

O padrão de replicação pode ser aplicado de forma análoga vis-à-vis também à pesquisa mais qualitativa. Um conjunto de dados de replicação de projetos qualitativos deve conter descrições minuciosas das regras de decisão seguidas, entrevistas conduzidas e informações coletadas. Transcrições de entrevistas, fotografias ou gravações de áudio podem facilmente ser digitalizados e incluídos em um conjunto de dados de replicação. Aderir ao padrão de replicação é mais difícil na pesquisa qualitativa e, por vezes, não pode ser seguido completamente. Mas justamente porque dados qualitativos detalhados são muito informativos e ricos, aderir ao padrão de replicação quando isso é possível vale a pena. Também seria útil para os pesquisadores qualitativos começar a discutir coletivamente as apropriadas aplicações ou modificações no padrão de replicação (ver Griffin e Ragin, 1994).

Uma vez que um conjunto de dados de replicação foi criado, ele deve ser disponibilizado publicamente e a referência a ele ser feita na publicação original (geralmente na primeira nota de rodapé). Uma abordagem é tornar as informações disponíveis sob pedido, mas isso pode ser inconveniente para ambos, tanto solicitantes quanto autores. Além disso, acadêmicos raramente são arquivistas profissionais. Seu comparativamente elevado grau de mobilidade também complica a distribuição própria, já que afiliações institucionais indicadas em artigos anteriores publicados não sejam mais corretas.

Esses problemas são resolvidos usando arquivos de dados profissionais. Arquivamento profissional implica backups de rotina online, duplicamento offline em um edifício ou cidade diferente, verificação contínua do meio digital em que o conjunto de dados está armazenado, o compromisso de armazenamento permanente (que envolve a alteração do meio de armazenamento conforme a tecnologia muda), publicidade frequente de forma a

que a contribuição do autor permanecerá amplamente conhecida, métodos rápidos e eficientes para distribuir esta informação a quem pede e financiamento suficiente para realizar cada uma dessas funções para o futuro indefinido.

Como observado, dois arquivos – o "Social Science Research Archive" do *Public Affairs Video Archive* (PAVA) na Universidade Purdue e o "Publication-Related Archive" do *Inter-University Consortium for Political and Social Research* (ICPSR) da Universidade de Michigan – irão agora aceitar bancos de dados de replicação. PAVA é o arquivo tecnicamente mais atualizado. Os funcionários irão tornar os dados disponíveis dentro de horas da submissão. Conjuntos de dados de replicação estão disponíveis instantaneamente via Internet por meio de servidores como "gopher", "FTP anônimo" e "Mosaic". Qualquer pessoa, em qualquer lugar do mundo, com uma conta de Internet tem acesso livre e ilimitado a esses dados.

O ICPSR é o mais antigo e mais conhecido dos dois arquivos. Sua equipe também irá manter e distribuir dados, e é atualmente capaz de distribuir dados relativos a publicações via FTP ou por e-mail a outros pesquisadores. O ICPSR também oferece outras classes para efetuar o depósito de dados. Se o autor da submissão estiver disposto a fornecer documentação adicional; para estes, a ICPSR irá fornecer vários níveis de verificação de dados e publicidade adicional. Desta forma, PAVA tem algumas vantagens tecnológicas sobre o ICPSR, mas o ICPSR ainda é a instituição mais conhecida e também oferece mais opções. Além disso, a apresentação de conjuntos de dados é gratuita e relativamente fácil em ambos os casos. Há pouco custo na submissão de dados para ambas as instituições (como é minha prática atual).

Conjuntos de dados de replicação podem ser submetidos a qualquer arquivo via disco ou fita, enviado por correio para PAVA e/ou ICPSR. Uma abordagem mais fácil e rápida é colocar os dados em um arquivo compactado auto-extraível e enviar os dados via FTP anônimo; o nome do arquivo, artigo, livro ou citação da dissertação, e um breve parágrafo descrevendo o conteúdo também devem ser incluídos em uma mensagem de e-mail que acompanha o arquivo.

Uma vez que um conjunto de dados de replicação é submetido e disponibilizado pelo arquivo, ele será anunciado por PAVA e ICPSR por meio de suas publicações regulares, catálogos e na Internet. Para maximizar a visibilidade, citações à publicação e correspondente banco de dados de replicação também aparecem em vários boletins informativos distribuídos aos membros da American Political Science Association (conforme descrito adiante).

### Comitês revisores de cargos e promoção

Comitês revisores de cargos e promoções julgam candidatos à promoção por suas contribuições para a comunidade acadêmica. A adesão ao padrão de replicação deve ser parte deste julgamento. Aqueles que seguem esse padrão são mais propensos a desfrutar de um público acadêmico mais amplo e têm a sua pesquisa mais bem compreendida e estendida; assim, eles vão ser levados mais a sério por seus pares acadêmicos. Além disso, os candidatos a cargos permanentes e promoção que submetem seus dados para um arquivo nacional devem ser reconhecidos por essa contribuição para a disciplina. Eu recomendo, por exemplo, que os pesquisadores adicionem uma seção extra ao seu *curriculum vitae* para "bancos de dados replicação arquivados." Esta importante contribuição seria, então, mais claramente notada, e a ela deve ser dado peso substancial em comissões, departamentos e decisões universitárias sobre promoção e estabilidade. No exterior do mundo acadêmico, articulistas também devem tomar nota destas contribuições significativas para a comunidade acadêmica.

### Programas de pós-graduação

O desenho de programas de pós-graduação e cursos específicos orientados para a pesquisa podem também incentivar a adesão ao padrão de replicação, que por sua vez pode fortalecer a capacidade dos alunos para aprender as noções básicas de pesquisa acadêmica e, finalmente, conduzir sua própria investigação original. A primeira "publicação" profissional para a maioria dos cientistas políticos é a dissertação do doutorado. Esta pretende ser uma contribuição original para o conhecimento e para a comunidade acadêmica. Para maximizar o impacto do trabalho de tese, os estudantes são aconselhados a apresentar bancos de dados de replicação para suas dissertações. Programas de pós-graduação podem também adotar regras que exigem que os estudantes apresentem os bancos de dados de replicação, quando apropriado. Ao fazê-lo, os programas de pós-graduação irão socializar e profissionalizar os estudantes nos padrões da disciplina.

PAVA aceita bancos de dados de replicação de dissertações e os embarga por um período de sua escolha. No Departamento de Governo da Universidade de Harvard, os estudantes fazendo pesquisa quantitativa e, quando aplicável, qualitativa, devem apresentar conjuntos de dados de replicação como um requisito para o grau de doutorado (é importante que o aluno crie e envie o banco de dados de replicação quando o trabalho está fresco em sua mente). Nós embargamos os dados por até cinco anos, conforme determinado pelo estudante e seu orientador, para dar ao estudante uma vantagem inicial na publicação. Na

maioria dos casos desde que nossa política tem sido adotada, os alunos optaram por um embargo curto ou mesmo nenhum.

Como observado, ter estudantes replicando os resultados de artigos existentes tem se provado ser uma ferramenta de ensino eficaz. Muitos programas de pós-graduação em economia até mesmo exigem que estudantes do doutorado reproduzam um artigo publicado em seu trabalho de segundo ano. Esta prática será útil para estudantes em ciência política quando conjuntos de dados de replicação se tornem mais amplamente disponíveis.<sup>6</sup>

### **Editores e pareceristas de revistas e livros**

Editores de revistas e editoras universitárias e comerciais trabalham duro para publicar produtos acadêmicos que tenham importantes contribuições para a disciplina da ciência política e um impacto máximo na profissão. Pelas razões descritas anteriormente, publicações de autores que aderem ao padrão de replicação são mais propensas a atender a esses critérios. Assim, os editores podem maximizar a influência de sua revista ou série de livros, exigindo adesão a este padrão.

Possivelmente, a abordagem mais simples é exigir dos autores que adicionem uma nota de rodapé em cada publicação indicando em qual arquivo público irá depositar a informação necessária para replicar seus resultados, além dos dados, quando isto estiver disponível. Para alguns autores, um comunicado explicando a inadequação desta regra, de prazos indeterminados de embargo dos dados ou parte dos dados, poderia substituir esta exigência. Neste caso, pareceristas seriam convidados a avaliar a declaração como parte do processo de avaliação geral e aconselhar o editor em conformidade com isto. Creio que devemos dar o máximo de flexibilidade para autores, respeitando seu direito de primeira publicação, a confidencialidade dos seus informantes ou por outras razões que serão discutidas a seguir. No entanto, essas exceções não se aplicam à grande maioria dos artigos e livros em ciência política.

Esta política é muito fácil de implementar, porque os editores ou suas equipes seriam responsáveis apenas pela existência da nota, não pela confirmação de que o conjunto de dados foi submetido nem verificar se os resultados realmente podem ser reproduzidos.

---

<sup>6</sup> Atualmente - antes do padrão de replicação ser amplamente adotado na disciplina – artigos baseados em replicação são frequentemente difíceis ou impossíveis. No entanto, outros procedimentos podem ser usados nesse ínterim por professores de métodos quantitativos. Por exemplo, eu tento ter estudantes apresentando um projeto de seus trabalhos finais no meio do semestre (geralmente com análises de dados, mas somente algumas páginas escritas), juntamente com um disco que contém um banco de dados de replicação. Estes bancos são, então, trocados aleatoriamente com outros alunos da turma. A tarefa da semana seguinte é repetir o projeto de seu colega de classe. Na maioria dos casos, o replicador e o autor original aprendem muito sobre os dados, métodos e processo de pesquisa.

Quaisquer verificações ou confirmações de questões da replicação podem e devem ser deixadas para futuros pesquisadores. Para conveniência dos editores e conselhos editoriais que estejam considerando adotar uma política como esta, o seguinte texto é um exemplo para essa política:

*Os autores dos artigos quantitativos publicados nesta revista [ou livros desta editora] devem indicar na sua primeira nota de rodapé em qual arquivo público irá depositar a informação necessária para reproduzir seus resultados quantitativos e a data em que serão submetidos. As informações depositadas devem incluir itens como dados originais, programas de computador especializados, listas de recodificação do software utilizado, extratos de arquivos de dados existentes, e um arquivo explicativo em que se descreva o que está incluso e explique como reproduzir os exatos resultados numéricos do trabalho publicado. Os autores podem encontrar no "Social Science Research Archive" do Public Affairs Video Archive (PAVA) da Universidade Purdue ou no "Publications-Related Archive" do Inter-University Consortium for Political and Social Research (ICPSR) da Universidade de Michigan lugares convenientes para depositar os seus dados. Declarações explicando a inadequação de partilha de dados para um trabalho específico (ou de prazos indeterminados de embargo dos dados ou partes destes) podem suprir esta exigência. Os pareceristas serão convidados a avaliar esta declaração como parte do processo de avaliação geral e aconselhar o editor em conformidade com isto. Os autores de obras baseadas em dados qualitativos são incentivados (mas não obrigados) a apresentar uma nota comparável que facilite a replicação sempre que possível. Como sempre, os autores são aconselhados a remover informações de seus conjuntos de dados que devem permanecer confidenciais, tais como os nomes dos entrevistados.*

Algumas revistas podem querer adotar exigências mais rigorosas (embora estas possam ser adequadas em determinadas circunstâncias, eu creio que geralmente não são necessárias ou desejáveis). Por exemplo, algumas revistas agora verificam se os dados foram, realmente, depositados em um arquivo público. Para os autores que solicitem embargos, algumas revistas podem querer exigir a submissão no momento da publicação e ter o arquivo embargado, de modo que o banco de dados de replicação seja preparado quando a mente do investigador está fresca. A seção de metodologia política da American Political Science Association (APSA) propôs um período de embargo máximo permitido de cinco anos. O Comitê de Estatísticas Nacionais, da National Academy of Science, recomenda ainda que os dados sejam disponibilizados ao público durante o processo de parecer.

Finalmente, algumas revistas podem querer experimentar pedindo a um revisor extra ou talvez um estudante de pós-graduação (atuando como editor-estagiário para a revista) que reproduza as análises para os artigos aceitos, mas ainda não publicados. Pareceristas de bancos de dados de replicação poderiam, então, fazer sugestões para os autores que poderiam

ser incorporadas antes da publicação, para tornar a replicação mais fácil ou mais clara para futuros pesquisadores. Esses tipos de experiências seriam muito úteis para revistas, autores, leitores e futuros pesquisadores. A exigência exata deve ser deixada para as necessidades de cada revista e editora, embora na ciência política a versão menos restritiva seja mais do que suficiente na implementação do padrão de replicação. Além disso, isto provavelmente se encaixaria melhor com as normas da disciplina.<sup>7</sup>

### **Exceções importantes**

Embora a presunção deva ser que os autores darão acesso livre aos dados de replicação, o editor, em acordo com o autor, terá sempre a palavra final sobre a aplicação de normas gerais para casos particulares. Exceções são essenciais quando as considerações de confidencialidade são importantes, para garantir direitos dos autores à primeira publicação e para uma variedade de outras razões. Embora estas exceções sejam importantes, elas provavelmente não serão aplicáveis à grande maioria dos trabalhos acadêmicos da área.

### **Confidencialidade**

Para manter a confidencialidade, organizações de pesquisa de opinião normalmente não divulgam os nomes e endereços dos entrevistados. Nestes e em casos afins, os autores confiam que tais informações possam cumprir com o requisito de replicação pela liberação de um subconjunto de seus dados, retiradas as informações de identificação. No entanto, em alguns casos, fornecer quaisquer dados seria inadequado. Por exemplo, em um artigo recente sobre as admissões de pós-graduação, meus co-autores e eu utilizamos dados do processo de admissões de Harvard (King, Bruce, e Gilligan, 1993). Claramente, as notas e o histórico acadêmico de alunos do programa não podem ser liberados. Na verdade, nós até retiramos os coeficientes de regressão do artigo, uma vez que senti que seria inadequado que futuros alunos fossem capazes de calcular as notas esperadas em nosso programa. A publicação de informações suficientes para permitir que alunos calculem a probabilidade de admissão também teria sido uma decisão impopular no gabinete do conselho geral de Harvard! No entanto, casos como esses são a exceção e não a regra.

---

<sup>7</sup> Outra ocasião em o padrão de replicação pode ser implementado é durante o processo de revisão por pares. Colaboradores de revistas e manuscritos de livros também devem verificar a existência de uma nota de rodapé indicando em que arquivo um banco de dados de replicação foi depositado. Uma vez que a nota de rodapé afeta a magnitude da contribuição do trabalho acadêmico, comentar sobre esta é provavelmente responsabilidade do parecerista. Além disso, sugerir aos autores que incluam esta nota de rodapé, e depositem os seus dados em um arquivo público, vai ajudar a lembrá-los e, talvez aos editores, deste método útil de contribuição acadêmica. Revistas também podem incluir pedidos para avaliar a nota de rodapé sobre replicação quando eles enviam seu pedido de revisão.

Em algumas situações raras, a confidencialidade não pode ser protegida se qualquer dado for disponibilizado ao público. Por exemplo, estudos baseados em entrevistas de elites entre uma população pequena, ou outros surveys com base em uma proporção muito grande da população relevante, potencialmente colocam este problema. Em um subconjunto desses casos, o autor pode ser capaz de tornar os dados disponíveis para os acadêmicos individuais dispostos a limitar a sua utilização de formas muito específicas (analogamente, regras específicas de análise de dados têm sido adotadas para os dados do Censo dos Estados Unidos, a fim de evitar a revelação de identidades individuais. Por exemplo, não são permitidos cruzamentos com menos de 15 indivíduos por célula). Estas são exceções importantes, mas elas também cobrem comparativamente poucos trabalhos publicados na ciência política.

Em algumas situações, os dados utilizados em um trabalho publicado não podem ser distribuídos porque são privados, tais como dados de survey do Roper Center. No entanto, a maioria destas organizações permite que os dados sejam redistribuídos pelos autores se forem modificados de alguma forma, tal como fazendo extratos das variáveis usadas ou recodificações. Dados totalmente de propriedade privada são raros na ciência política.

### Direitos à primeira publicação

Como indicado anteriormente, para garantir o direito de primeira publicação, é adequado submeter os dados a um arquivo público e solicitar um embargo por algum período de tempo específico. No entanto, como estes embargos devem ser relativamente raros, o padrão de replicação exige que sejam fornecidos apenas os dados efetivamente utilizados em uma publicação. Por exemplo, se você conduziu o seu próprio survey com 300 variáveis e usou somente 10 para um artigo, você precisará fornecer apenas aquelas 10. Se você tiver uma variável com cinco categorias e utilizar apenas duas das categorias no artigo, você poderia fornecer a variável recodificada com apenas duas categorias. Se você tem 1500 observações, e usar apenas 1000 delas no artigo (talvez retirando os estados do sul, por exemplo), você também precisa apresentar apenas os 1000 casos usados em sua análise. Então você pode guardar as informações restantes para suas futuras publicações. Você certamente poderia fornecer o resto dos dados, o que provavelmente faria o seu trabalho mais valioso para a comunidade acadêmica, mas a decisão de fazê-lo seria sua.

Em alguns casos, os autores podem desejar o embargo do subconjunto de dados utilizado em um artigo para arrumar, documentar, e, então, publicizar um conjunto de dados maior de onde foi extraído. Isso seria mais conveniente para o investigador, e pode também

beneficiar futuros pesquisadores, incentivando-os a esperar pela versão mais abrangente dos dados (nestes casos, os investigadores devem manter uma versão antiga dos dados, ou documentar totalmente quaisquer alterações nos dados desde a publicação do artigo).

Em termos gerais, o ponto básico dessa proposta é alterar as expectativas dos autores, da atual situação de raramente tomar qualquer medida para garantir que o seu trabalho possa ser replicado, para geralmente dar alguns passos nesta direção. As exceções são importantes, mas não se aplicariam à grande maioria dos artigos publicados.

### Apoio a estas políticas

#### Adoção de política de replicação

O apoio formal a estas políticas parece estar crescendo. Desde o ano passado, a *American Journal of Political Science*, sob a direção de Kenneth Meier, e a *Political Analysis* (a revista de métodos da área), sob a edição de John Freeman, agora exigem que notas de rodapé sobre conjuntos de dados de replicação sejam incluídas em todos os artigos. Os editores não têm encontrado nenhuma resistência dos autores, e a política tem exigido muito pouco tempo e esforço para ser implementada (Kenneth Meier também relata que 70% dos artigos empíricos que aceitou utiliza coleções de dados originais coletados pelo autor). A *British Journal of Political Science* (David Sanders e Albert Weale, editores), e a *Policy Studies Journal* (editado por Uday Desai e Mack C. Shelley II) adotaram políticas semelhantes, e a *International Interactions*, sob edição de Harvey Starr, está no processo de fazê-lo.

A nova política da editora de ciência política e direito da Universidade de Michigan é "esperar rotineiramente que todos os autores tenham uma nota de rodapé em seus livros, indicando onde o seu banco de dados de replicação está arquivado (e quando ele estará disponível, se for o caso disso)", embora não seja um requisito absoluto para publicação. A série da Cambridge University Press "Economia política das instituições e decisões", sob a editoria de James E. Alt e Douglass North, adotou recentemente uma versão desta política. A Free Press e a HarperCollins fizeram o mesmo. Muitos outros editores e conselhos editoriais na ciência política indicaram apoio a uma política de replicação, mas ainda estão no processo de consideração da forma específica que a política adotará.

A política dos programas de ciência política na National Science Foundation (NSF) em relação a conjuntos de dados de replicação é claramente indicada nas suas cartas de premiação: "Todos os bancos de dados produzidos com o auxílio deste prêmio devem ser arquivados em uma biblioteca de dados aprovada pelo responsável pelo programa, o mais tardar um ano após o vencimento da bolsa." Para aplicar esta regra, os programas de ciência

política da NSF adotaram recentemente diversas novas políticas. Em primeiro lugar, todas as publicações em revistas e livros que foram preparados com a ajuda financeira da NSF devem incluir uma declaração indicando em qual arquivo público que será depositada a informação necessária para reproduzir seus resultados e os dados que serão submetidos (ou uma explicação se não puderem ser). Em segundo lugar, quando for concedida uma bolsa, o responsável do programa vai pedir ao potencial pesquisador que verifique se ele ou ela alocou fundos suficientes para cumprir a exigência de arquivamento de dados. Em terceiro lugar, dentro de um ano a partir da data de encerramento de um subsídio, os investigadores devem informar o responsável do programa de ciência política onde seus dados foram depositados. Finalmente, os responsáveis dos programas na NSF irão considerar a conformidade com esta política de arquivamento de dados um importante critério adicional para julgar os candidatos a renovações e novas verbas.

Qualquer um que receba fundos do National Institute of Justice (NIJ) agora deve "entregar para o NIJ, após a conclusão do projeto, cópias legíveis por computador e documentação adequada de todas as bases de dados e programas desenvolvidos ou adquiridos relacionados à pesquisa" (Sieber 1991, 9). O Comitê de Estatísticas Nacionais, da National Academy of Sciences, também recomendou políticas semelhantes às sugeridas aqui (ver Fienberg et al., 1985). Embora existam muitas diferenças nacionais em normas de compartilhamento de dados, políticas e recomendações relacionadas, estas foram adotadas ou abordadas por governos nacionais, organizações internacionais, acadêmicos e sociedades profissionais, agências financiadoras, outras disciplinas e revistas acadêmicas (ver Boruch e Cordray, 1985).

Para ajudar a dar visibilidade adicional a autores de bancos de dados de replicação, *The Political Methodologist*, newsletter da seção de metodologia política da APSA (editada por Michael R. Alvarez e Nathaniel Beck), anunciou que referências de banco de dados, uma citação para o artigo associado e um breve resumo serão destacados nas edições futuras. Citações similares aparecerão em outros boletins informativos, se os dados forem relevantes para seus focos temáticos; estes boletins incluem *Policy Currents* (editado por Laura Brown), *Law and Courts* (editada por Lee Epstein), *Urban Politics Newsletter* (Arnold Vedlitz, editor), *Computer and Multimedia Section Newsletter* (editada por Bob Brookshire), *Caucus for a New Political Science Newsletter* (John C. Berg, editor), *Clio: The Newsletter of Politics and History* (Dave Robertson, editor), e *VOX POP*, newsletter da seção de partidos políticos e grupos de interesse da APSA (editada por John Green).

## Discussões sobre políticas de replicação

Políticas de replicação têm sido amplamente discutidas na comunidade da ciência política nos últimos anos. Entre metodólogos políticos, o apoio é entusiasmado e parece ser unânime. Longos debates formais e informais foram realizados nas últimas três reuniões anuais de verão do *Political Methodology Group* (Universidade de Wisconsin, Madison, 1994; Universidade da Flórida, 1993; e Universidade de Harvard, 1992). Painéis bem frequentados nas duas últimas reuniões da *American Political Science Association* também têm se dedicado em parte (em 1993) ou na íntegra (em 1994) a esta questão. A seção de metodologia política da APSA aprovou por unanimidade uma resolução em 1994 pedindo a todos os editores de revistas na disciplina que exijam notas de rodapé indicando onde os bancos de dados de replicação estão armazenados.

A seção de política comparada da APSA realizou um debate sobre esta questão na convenção anual de 1994. Após um debate considerável na convenção enfocando as preocupações especiais que comparatistas têm sobre confidencialidade e distribuição de dados "descontextualizados", o comitê executivo da seção endossou a ideia em termos gerais. O comitê escreveu posteriormente uma declaração de política proposta que reflete as preocupações especiais dos comparativistas, ao mesmo tempo em que exige a nota de rodapé sobre replicação. Esta proposta está agora sendo distribuída por meio do boletim informativo da seção para um comentário geral de seus membros.

Discussões abrangentes também foram realizadas em reuniões de muitos dos conselhos editoriais da nossa disciplina e em reuniões de seção da *American Political Science Association* e, a julgar pela resposta nestes fóruns, o suporte ao padrão de replicação é forte em toda a disciplina. Muitas perguntas perspicazes e questões foram levantadas sobre os métodos específicos para a implementação de políticas de replicação. As propostas discutidas neste artigo foram melhoradas como resultado.

## Questões e respostas

No decurso de inúmeras conversas sobre estes tópicos, várias questões foram levantadas e discutidas. Eu listo algumas destas aqui, junto com as resoluções mais comuns.

*Será que um padrão de replicação irá reduzir os incentivos para que os investigadores coletem conjuntos de dados grandes e difíceis?*

Investigadores recebem o reconhecimento pela coleta de dados e por torná-los disponíveis para a comunidade acadêmica. Este reconhecimento ocorre na forma de citações

aos dados e aos artigos do autor, agradecimentos pela ajuda do autor, e promoção, estabilidade e aumento salarial. O trabalho acadêmico é julgado por sua contribuição, portanto, tornar um artigo mais importante, contribuindo com um conjunto de dados de replicação só pode melhorar o reconhecimento que o autor recebe. O risco é não ter as próprias ideias roubadas, um medo familiar, mas em grande parte infundado, que a maioria de nós têm experimentado ao escrever dissertações; um risco muito maior - de fato um risco com uma alta probabilidade - é ter uma de suas publicações ignorada. Submeter um conjunto de dados de replicação pode diminuir significativamente essa probabilidade.

Além disso, como discutido antes, as informações coletadas mas não utilizadas no artigo não necessitam de serem incluídas no banco de dados de replicação; apenas as variáveis e observações necessárias para repetir os resultados publicados precisam ser apresentadas. Se acontecer de existir novas informações nas variáveis e observações apresentadas, o autor terá uma vantagem substancial em extrair tais informações. Na maioria dos casos, cerca de dois anos decorrem entre a realização de um artigo e a publicação final. Se este não é tempo suficiente em uma determinada ocasião, o autor ainda pode submeter o conjunto de dados e escolher um embargo por um período determinado, ou mesmo comprometer-se a enviá-lo em uma data futura especificada.

*Implementar o padrão de replicação vai fazer com que muito mais dados estejam disponíveis em arquivos públicos. Não será uma consequência não intencional da proposta que os futuros pesquisadores vão gastar a maior parte de seu tempo analisando os dados existentes, em vez de coletar novos dados, gastando mais tempo no laboratório de informática, em vez de no campo?*

A experiência sugere exatamente o oposto. Quando o ICPSR foi fundado, e mais tarde expandido, a quantidade de dados disponíveis publicamente aumentou dramaticamente. No entanto, análises de conteúdo indicam que muitos mais artigos que contém dados originais foram publicados durante este tempo (King, 1991). Assim, parece que o aumento da disponibilidade de dados originais inspira outros estudiosos a coletar seus próprios dados originais. Na verdade, se aprende tanto reproduzindo pesquisas existentes que torna a coleta de novos dados, seguindo procedimentos bem-sucedidos desenvolvidos em pesquisas anteriores, muito mais fácil.

*Não seria melhor se todas revistas e séries de livros adotassem exatamente a mesma política ao mesmo tempo?*

Pode haver algumas vantagens na coordenação, mas a razão pela qual temos diferentes revistas à frente argumenta contra a espera. Cada revista tem um público diferente, segue diferentes manuais de estilo, tem diferentes padrões de qualidade, diferentes conselhos editoriais, diferentes editores, diferentes pareceristas, diferentes estilos metodológicos, diferentes editores de texto, e incentiva uma diferente mistura de artigos. Não deveria ser surpreendente ou preocupante se diferentes periódicos adotem políticas ligeiramente diferentes em relação à replicação, ou as adotem de acordo com diferentes calendários.

*Se nossa revista exige a adesão ao padrão de replicação, os autores não vão enviar seus trabalhos para outro lugar ou não publicar artigos e guardar o seu trabalho até que um manuscrito de livro esteja pronto?*

Isso pode ser verdade para alguns autores, mas não tem sido a experiência das revistas que já adotaram esta política. Além disso, muitas editoras de livros estão adotando a mesma política, e ninguém pode recordar um pesquisador que tenha recusado verbas da NSF a fim de evitar esta regra. Uma vez que a política de replicação esteja devidamente comunicada e explicada aos autores, provavelmente eles vão entender que é do seu interesse. Também é claramente do interesse de revistas ter seus artigos citados e, portanto, seguir o padrão de replicação. Além disso, não seguir esta norma seria muito mais injusto com os potenciais leitores, e mais prejudicial para a profissão.

*Se eu lhe der os meus dados, não há uma chance de você descobrir que eu estou errado e divulgar isto a todos?*

Sim. A maneira pela qual a ciência avança é nos fazendo vulneráveis a estar errados. Em última instância, todos nós estamos buscando um objetivo comum: uma compreensão mais profunda do governo e da política. Assim, devemos dar aos outros a oportunidade de provar que estamos errados. Apesar de ser criticado nem sempre é agradável, é prova inequívoca de que está sendo levado a sério e fazendo alguma diferença. Novamente, ser ignorado - o destino de mais de 90% de todas as publicações de ciência política - é um risco muito mais sério.

*Não devem os editores recolher conjuntos de dados de replicação para garantir que tenham sido submetidos?*

Esta é uma possibilidade, mas os editores não são arquivistas mais profissionais do que os autores. Editores poderiam muito bem aproveitar-se de arquivos como PAVA ou

ICPSR. Eu também não acho que a verificação seja necessária, uma vez que quaisquer discrepâncias serão corrigidas pelos pesquisadores futuros.

*Se todo mundo começar a submeter seus bancos de dados de replicação, os arquivos não vão rapidamente começar a ficar cheios de lixo?*

Isto é extremamente improvável. Diferentemente da maioria dos bancos de dados submetidos aos arquivos públicos, conjuntos de dados de replicação terão sido filtrados pelo processo de revisão por pares, e provavelmente serão verificados por futuros pesquisadores. Assim, arquivos públicos podem aproveitar o processo científico regular para construir valor para suas coleções. Além disso, o tamanho médio de um banco de dados na ciência política é menor do que um megabyte. Mesmo se cerca de 100 bancos de dados de replicação forem submetidos a um arquivo a cada ano, cerca de US\$ 600 dólares para um gigabyte de espaço em disco e uma conexão com a Internet facilmente acomodaria todos os envios por bem mais de uma década.

*Se a submissão de conjuntos de dados de replicação é do interesse de pesquisadores individuais, por que precisamos que revistas e editoras exijam sua apresentação?*

Nós não precisamos de regras quando este é o costume, mas a experiência em nossa disciplina e diversas outras indica que o problema dos bens coletivos não pode ser resolvido, na prática, sem alguma mudança de políticas. Veja Dewaid, Thursby, e Anderson (1985), Boruch e Cordray (1985; 209-210), e Fienberg et al. (1985) para justificativas mais detalhadas.

*Por que estamos nos preocupando com o que poderia ser chamado de "duplicação" da investigação já existente? Não é mais importante a questão da replicação real, onde as mesmas medidas são aplicadas a novas áreas, países ou períodos de tempo?*

Boa ciência exige que sejamos capazes de reproduzir resultados existentes, e que outros pesquisadores sejam capazes de mostrar como achados substantivos mudam na medida em que aplicamos os mesmos métodos em novos contextos. A última etapa é mais interessante, mas isso não reduz a necessidade da anterior. Na verdade, podemos encorajar os acadêmicos a prosseguir a replicação em novos contextos, se eles puderem ter mais certeza sobre os resultados presentes. Um melhor conhecimento dos resultados existentes, por meio da distribuição dos conjuntos de dados de replicação, também permitirá uma adaptação mais fácil dos métodos e procedimentos existentes a novos contextos. Além disto, uma prática

usual na estimativa de efeitos causais é realizar uma mudança de cada vez, para que possamos relacionar mudanças individuais a efeitos específicos e julgar cada efeito de modo isolado.

## Referências

- Boruch, Robert F., and David S. Cordray. 1985. "Professional Codes and Guidelines in Data Sharing." In *Sharing Research Data*, Stephen E. Fienberg, Margaret E. Martin, and Miron L. Straf, eds., 199-223. Washington, DC: National Academy Press.
- Ceci, Stephen, and Elaine Walker. 1983. "Private Archives and Public Needs," *American Psychologist* 38 (April):414-23.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. 1986. "Replication in Empirical Economics: The Journal of Money, Credit, and Banking Project," *American Economic Review* 76, 4 (September):587-603.
- Diamond, A. M., Jr. 1988. "The Polywater Episode and the Appraisal of Theories." In *Scrutinizing Science*, A. Donovan et al., eds. New York: Kluwer Academic Publishers.
- Feigenbaum, Susan, and David M. Levy. 1993. "The Market for (Ir)reproducible Econometrics (with comments and response)," *Social Epistemology* 7, 3 (July-September):215-92.
- Fienberg, Stephen E., Margaret E. Martin, and Miron L. Straf, eds., *Sharing Research Data*. Washington, DC: National Academy Press.
- Griffin, Larry, and Charles C. Ragin. 1994. "Formal Methods of Qualitative Analysis," a special issue of *Sociological Methods and Research* 23, 1.
- Hamilton, David P. 1991. "Research Papers: Who's Uncited Now?" *Science*, 4 January, p. 25.
- King, Gary. 1991. "On Political Methodology," *Political Analysis* 2: pp. 1-30.
- King, Gary, John M. Bruce, and Michael Gilligan. 1993. "The Science of Political Science Graduate Admissions," PS: *Political Science & Politics* 4 (December):772-78.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Leimer, D. R., and Lesnoy, S. D. 1982. "Social Security and Private Saving," *Journal of Political Economy* 90:60&29.
- Neuliep, James W., ed. 1991. Replication Research in the Social Sciences. Newbury Park, CA: Sage Publications. Replication, Verification, Secondary Analysis, and Data Collection in Political Science Paul S. Herrnson,' University of Maryland, College Park

KING, Gary. *Replicação, replicação.*

Pendlebury, David. 1994. (Institute for Scientific Information.) Telephone conversation. 8 September.

Sieber, Joan E., ed., 1991. *Sharing Social Science Data: Advantages and Challenges*, Newbury Park, CA: Sage Publications.