

## Nem melhor nem pior, apenas diferente

Neither better nor worse, simply different

Ni mejor ni peor, sólo diferente

Cláudia Medina Coeli <sup>1</sup>  
 Rejane Sobrino Pinheiro <sup>1</sup>  
 Marília Sá Carvalho <sup>2</sup>

<sup>1</sup> Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

<sup>2</sup> Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

### Correspondência

C. M. Coeli  
 Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro.  
 Pça. Jorge Moreira Machado  
 1000, Cidade Universitária,  
 Rio de Janeiro, RJ  
 21941-598, Brasil.  
 coelicm@gmail.com

Alguma vez você se sentiu discriminada(o) por empregar dados secundários em sua pesquisa? Sendo as três autoras que assinam este texto pesquisadoras que têm no desenvolvimento e aplicação de técnicas para o uso de dados secundários a principal linha de pesquisa, a resposta é definitivamente não. Entretanto, não é raro ouvirmos relatos de colegas que encontraram alguma barreira para o desenvolvimento de teses, ou conseguir financiamento em função da opção pelo uso de dados secundários.

Em artigo recente, Rothman <sup>1</sup> abordou seis percepções equivocadas sobre aspectos da pesquisa epidemiológica que costumam ser reforçadas em salas de aula e livros texto. Embora o autor não tenha abordado as fontes de dados, acreditamos que um sétimo erro deveria ser adicionado à lista: a noção de que dados primários são a única fonte válida para a condução de estudos epidemiológicos.

Dados populacionais, vitais, epidemiológicos, administrativos e clínicos sofreram mudanças expressivas na forma como são produzidos e disseminados. São hoje disponibilizados em bases eletrônicas que contemplam milhões de microdados individuais. Além das fontes tradicionais antes apontadas, surgiram outras modalidades. As marcas digitais produzidas no acesso a diferentes plataformas de comunicação na Internet e no uso de celulares, passaram a ser

empregadas em estudos sobre a influência de padrões de comportamentos e mobilidade na determinação e disseminação de doenças <sup>2</sup>.

Dados secundários possuem potencial para apoiar estudos sobre questões de grande relevância à Saúde Coletiva, acentuados pela ampla disponibilidade, abrangência e cobertura. Na verdade, são o melhor dado para responder a perguntas sobre os determinantes das taxas de incidência nas populações, como sugerido por Rose <sup>3</sup>. Ainda assim, cabe discutir a integração dos dois mundos. A interação gene-ambiente, por exemplo, demanda o uso de populações de estudo com tamanhos cada vez maiores. O contexto da *big epidemiology* <sup>4</sup> estimula a prática do *data sharing*, fazendo com que dados coletados para estudos específicos passem a ser utilizados por pesquisadores não originalmente envolvidos no seu planejamento e execução.

A era do *big data* tem levado à recomendação do uso deste rico acervo na pesquisa <sup>5</sup>, incluindo a pesquisa populacional em saúde <sup>6</sup>. Alguns autores apontam, entretanto, a necessidade do uso responsável dessas bases de dados <sup>7</sup>. As principais críticas feitas às fontes de dados secundários são a ausência de mecanismos para a garantia e controle da qualidade dos dados coletados, e a falta de variáveis que seriam necessárias para testar adequadamente hipóteses causais no nível do indivíduo.

A questão da qualidade é fundamental. Por um lado, deve-se proceder à avaliação de diferentes dimensões de qualidade <sup>8</sup> previamente ao uso de uma fonte de dados secundários. Por outro, os responsáveis pela custódia de bases devem empregar técnicas para prevenção, detecção e reparo de erros <sup>9</sup>, e disponibilizar farta documentação sobre seus acervos. O financiamento de infraestrutura para a manutenção e acesso de dados é um elemento essencial que deve ser considerado nas políticas de estímulo ao uso de dados secundários <sup>5,6</sup>. Com relação às variáveis disponíveis para análise, a integração de bases de dados, por meio do emprego de técnicas de *linkage* de registros <sup>10</sup>, pode contribuir para uma melhor especificação de variáveis de exposição e desfecho, além de ampliar o número de variáveis disponíveis para o ajuste de fatores de confusão. Adicionalmente, algumas soluções metodológicas têm sido propostas para mitigar o problema de fatores de confusão não mensurados <sup>11</sup>. Por fim, cresce o interesse em responder a perguntas de pesquisa não etiológicas, para as quais não é necessário o ajuste para confundimento. Esse é o caso de perguntas voltadas para a avaliação de intervenções de Saúde Pública, que podem ser respondidas empregando dados de natureza diversa, juntamente com a aplicação de novas técnicas analíticas como, por exemplo, a mineração de dados e a modelagem computacional de sistemas complexos <sup>2,6,10</sup>.

Para além das questões metodológicas, o uso responsável também deve contemplar os aspectos de respeito à privacidade. É necessário o desenvolvimento de um referencial ético que leve em consideração as particularidades da pesquisa baseada em dados secundários, especialmente no que diz respeito ao consentimento informado <sup>12</sup>. No Brasil, recentemente, foi sancionada a *Lei nº 12.527*, que regulamenta o acesso à informação pública <sup>13</sup>. Deve-se tomar cuidado para que as interpretações conservadoras da lei não impliquem restrições desnecessárias à divulgação de conteúdos nas bases não identificadas,

ou ao acesso, com as salvaguardas necessárias e para fins de pesquisa, às bases identificadas. Um estudo promovido pelo Instituto de Medicina dos Estados Unidos apontou que a legislação americana, que aborda a transferência de informações em saúde (*HIPAA Privacy Rule*), teve impactos negativos sobre a condução de pesquisas relevantes para a Saúde Pública <sup>14</sup>. É preciso buscar um equilíbrio entre os direitos individuais e os interesses coletivos, de forma a não inviabilizar, por exemplo, estudos que visem a melhorar a saúde, a assistência e a vida dos usuários do SUS.

O uso de dados secundários na pesquisa demanda investimentos na formação de recursos humanos. Se por um lado cada vez mais é necessária a incorporação de profissionais da área de tecnologia de informação às equipes de pesquisa, é preciso formar pesquisadores da área de Saúde Coletiva que sejam capazes de interagir com estes profissionais, como *experts* interativos, tal como definido por Collins et al. <sup>15</sup>. Que habilidades devem ser desenvolvidas e o nível de expertise mínimo esperado são questões a serem definidas. SQL (*Structured Query Language*), *record linkage*, integração de dados não estruturados, mineração de dados e modelagem computacional de sistemas complexos são alguns conteúdos que deveriam ser abordados.

Este texto foi finalizado no Rio de Janeiro durante o período do Carnaval de 2014, que tem no desfile das escolas de samba um dos eventos turísticos mais importantes da cidade. Seu título é inspirado em um bordão criado nos anos 1960 por Nelson de Andrade, então presidente do Grêmio Recreativo Acadêmicos do Salgueiro. O bordão original *Nem melhor, nem pior, apenas uma Escola diferente* buscava enfatizar a revolução criativa no carnaval carioca promovida por Fernando Pamplona e Arlindo Rodrigues <sup>16</sup>. Dados secundários representam uma fonte valiosa para a pesquisa em Saúde Coletiva. Tirar o máximo proveito de seu uso também demanda uma revolução. É preciso pensar diferente, formar diferente e fazer diferente.

## Colaboradores

Todas as autoras participaram de todas as etapas de produção do texto.

1. Rothman KJ. Six persistent research misconceptions. *J Gen Intern Med* 2014; Epub ahead of print.
2. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. *PLoS Comput Biol* 2012; 8:e1002616.
3. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985; 14:32-8.
4. Thompson A. Thinking big: large-scale collaborative research in observational epidemiology. *Eur J Epidemiol* 2009; 24:727-31.
5. Community cleverness required. *Nature* 2008; 455:1.
6. Mabry PL. Making sense of the data explosion: the promise of systems science. *Am J Prev Med* 2011; 40(5 Suppl 2):S159-61.
7. Hernán MA. With great data comes great responsibility. *Epidemiology* 2011; 22:290-1.
8. Lima CRA, Schramm JMA, Coeli CM, Silva MEM. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cad Saúde Pública* 2009; 25:2095-109.
9. Herzog TN, Scheuren FJ, Winkler WE. *Data quality and record linkage techniques*. New York: Springer; 2007.
10. Christen P. *Data matching concepts and techniques for record linkage, entity resolution, and duplicate detection*. Berlin/New York: Springer; 2012.
11. Toh S, García-Rodríguez LA, Hernán MA. Analyzing partially missing confounder information in comparative effectiveness and safety research of therapeutics. *Pharmacoepidemiol Drug Saf* 2012; 21:13-20.
12. da Silva MEM, Coeli CM, Ventura M, Palacios M, Magnanini MM, Camargo TM, et al. Informed consent for record linkage: a systematic review. *J Med Ethics* 2012; 38:639-42.
13. Ventura M. Lei de acesso à informação, privacidade e a pesquisa em saúde. *Cad Saúde Pública* 2013; 29:636-8.
14. National Research Council. *Beyond the HIPAA privacy rule: enhancing privacy, improving health through research*. Washington DC: The National Academies Press; 2009.
15. Collins H, Evans R, Ribeiro R, Hall M. Experiments with interactional expertise. *Stud Hist Philos Sci Part A* 2006; 37:656-74.
16. Faria GJM. Nem melhor, nem pior, apenas uma escola diferente: os Acadêmicos do Salgueiro e as transformações estéticas e ideológicas na cultura brasileira (1959-1971). *Revista Litteris* 2010; (6). <http://www.revistaliteris.com.br>.

Recebido em 14/Abr/2014

Aprovado em 15/Abr/2014