

Compartilhamento de dados de pesquisa: para ser ético, equitativo e responsável há um grande investimento a fazer

Alúcio J D Barros, PhD*

Bernardo L Horta, PhD*

* Centro de Pesquisas Epidemiológicas, Universidade Federal de Pelotas, Pelotas, Brasil

Setembro de 2018

Muito se tem discutido sobre o compartilhamento de dados nos últimos anos, sendo que alguns periódicos já adotaram uma política compulsória de disponibilização de dados em repositórios públicos, a exemplo dos periódicos PLOS (como PLOS One, PLOS Medicine). Neste comentário, trazemos a discussão para o contexto brasileiro da área de saúde coletiva.

A ideia de compartilhar dados de forma ampla surgiu com a experiência do projeto genoma e dados genômicos passaram a ser compartilhados rotineiramente.¹ Mas rapidamente a prática passou a ser utilizada em estudos laboratoriais,² nos quais se usam animais ou células, por exemplo, como unidades de análise, e os desenhos são focados em responder uma questão de pesquisa bastante objetiva. O compartilhamento de estudos sobre saúde humana, em nível individual, é bastante mais complicado em termos éticos³ e em termos da estrutura e quantidade de dados.

No âmbito de inquéritos nacionais sobre saúde, nutrição ou condições de vida, financiados com recursos governamentais, disponibilizar os dados de forma pública e irrestrita tem sido a regra no Brasil e na maioria dos países já há muitos anos. Não só não se questiona a prática, como, na verdade, se espera que estes dados sejam colocados à disposição da comunidade para que possam ser utilizados ao máximo. Casos em que isso não acontece, como o conjunto de inquéritos STEPS¹, iniciativa da Organização Mundial da Saúde em parceria com os governos dos países causam surpresa.³ A razão principal para que os dados sejam disponibilizados livremente é maximizar seu uso e seu valor, respeitando os interesses da comunidade ou grupo que fornece os dados e também de quem paga, direta ou indiretamente, pela pesquisa.

¹ STEPwise approach to surveillance (<http://www.who.int/ncds/surveillance/steps/reports/en/>).

Com relação a projetos de pesquisa conduzidos por grupos ou indivíduos, a justificativa que se tem dado para defender a disponibilização dos dados de forma pública segue linha de raciocínio semelhante – maximizar a utilização dos dados e devolver mais benefícios para a população que participou do estudo ou o financiou através de impostos.⁴ E também se quer evitar que os dados sejam subutilizados.¹ Mas, nesses casos, a propriedade intelectual do projeto e dos dados é mais claramente definida como dos investigadores principais que elaboraram o projeto e asseguraram o seu financiamento e execução.

Uma revisão ampla sobre o tema foi preparada por Susan Bull, em relatório divulgado pelo *Wellcome Trust* em 2016.² A seguir, apresentamos os principais pontos que têm sido alvo de discussão, antes de entrar nos aspectos específicos de nossa área. Inicialmente, se tem dado ênfase a um processo de compartilhamento de dados *equitativo*, que respeita os interesses e os direitos de todas as partes envolvidas. Um dos temores citados na discussão do tema está relacionado a um possível desestímulo daquele que faz a pesquisa primária, e investiu enorme quantidade de tempo e energia em escrever um projeto, conseguir financiamento e coletar os dados, em função de compartilhar os seus dados com quem nada contribuiu para o projeto.^{2,3} Outra preocupação é que publicações independentes, sem colaboração com os investigadores originais, por grupos melhor equipados para realizar análises e melhor relacionados acabem por ser mais relevantes, mais citadas, mais visíveis.²

Outra questão central é a necessidade de “curar” os dados antes de disponibilizá-los, ou seja, preparar um conjunto de materiais que descrevam completamente a pesquisa, incluindo instrumentos e estratégias de coleta, documentação, limpeza e codificação dos dados.⁵ No caso brasileiro, uma necessidade adicional será a preparação de documentação em inglês para um compartilhamento em nível internacional, incluindo a tradução dos instrumentos e questionários. Isso exige capacidade técnica e recursos financeiros. Importante notar que o compartilhamento de dados mal documentados pode levar a críticas de outros grupos, com prejuízo à reputação de quem os gerou, além da produção de resultados equivocados ou enviesados. E há também a preocupação com a utilização dos dados sem que seja dado o devido crédito a quem gerou os dados.²

Finalmente, mas não menos importante, há a necessidade de que se garanta a privacidade e o anonimato dos participantes, assim como se evite que grupos ou comunidades sejam estigmatizados em função da publicação dos dados. Pearce e Smith³ discutem essa questão com

grande ênfase e, embora concordem que dados de estudos epidemiológicos devem ser disponibilizados para reanálise e reinterpretação, consideram que a exigência incondicional de publicização de dados após o término de um estudo é antiética.

Feito este breve resumo, discutimos o compartilhamento de dados em nosso contexto não como filósofos, mas como pesquisadores com larga experiência no desenho de projetos, coleta de dados, análises e profundo conhecimento da área.

O compartilhamento equitativo e responsável de dados de estudos em geral, e de estudos epidemiológicos, em especial, tem custo elevado, pois exige cuidadosa preparação de documentação e das informações que permitam, em princípio, o uso adequado dos dados por terceiros sem a intervenção do pesquisador primário, que gerou o dado. O compartilhamento de dados, em vista dos custos envolvidos, tem que estar previsto e orçamentado desde a fase de planejamento do estudo para que possa acontecer de forma adequada.

A questão do financiamento é importante visto o cenário nacional de fomento à pesquisa, limitado a poucas agências com orçamentos variáveis e atualmente com tendência a decrescer. Veja que ainda não resolvemos a questão do custo de publicação de artigos em acesso aberto, prática considerada desejável pelas agências e pelos investigadores. A postura das nossas agências de financiamento tem sido errática sobre o assunto, ora permitindo, ora não, que o custo de publicação seja incluído nos projetos de pesquisa. É irresponsável jogar sobre o investigador a responsabilidade de compartilhar dados sem correspondente apoio financeiro.

A disponibilidade orçamentária, somente, não resolve o problema. Há também a necessidade de criação de uma competência na área, assim como orientação de boas práticas. Em nosso país é difícil formar equipes multidisciplinares e a contratação de pessoal especializado em tratamento de dados é uma barreira a ser considerada nas instituições públicas, visto que não se pode contratar pessoal temporário, há necessidade de abertura de vaga, concurso, etc., processo que leva tempo e muita articulação política. Em um momento em que a contratação de docentes está limitada e o sistema de pós-graduação sob grave ameaça (vide ameaça recente de corte orçamentário drástico da Capes), pensar em contratar pessoal especializado para tratamento de dados parece uma realidade distante.

O segundo ponto que queremos abordar está ligado aos tipos de estudo realizados mais frequentemente em nossa área. Em vista do tema, vamos nos ater a estudos baseados majoritariamente em instrumentos fechados e pré-codificados. Estudos essencialmente

qualitativos baseados em estratégias como grupos focais e entrevistas em profundidade apresentam dificuldades bastante diferentes e estão menos sujeitos, pelo menos por enquanto, à pressão de periódicos e financiadores para que os dados sejam compartilhados de forma pública.

Os estudos mais comuns são inquéritos transversais, mas também temos estudos longitudinais e ensaios clínicos e comunitários, entre outros. Por simplicidade, vamos chamar esse conjunto de estudos produzidos pela área de estudos epidemiológicos. A publicização desses dados envolve, em primeiro lugar, enorme cuidado ético. De rotina, o pesquisador se compromete com os participantes a realizar apenas análises em nível de grupos, mantendo sempre o anonimato e a privacidade dos participantes. A disponibilização dos dados de forma pública exige que o pesquisador garanta que os dados sejam não identificáveis, o que nem sempre é evidente. Em casos onde o anonimato é crítico e difícil de ser garantido por métodos tradicionais, pode-se utilizar de técnicas de imputação múltipla para gerar bancos de dados equivalentes em termos de informação, mas que não permitem a identificação dos participantes.⁶ Note que existe a necessidade de pessoal capacitado para garantir o anonimato, especialmente em estudos mais localizados, como é o caso de coortes de nascimento que incluem nascidos em uma determinada cidade, num determinado ano, por exemplo.

Diferentemente de estudos laboratoriais, estudos epidemiológicos servem para uma ampla gama de análises, que podem levar algum tempo a serem realizadas. Desta forma, a definição de períodos de embargo dos dados para que os pesquisadores possam realizar as análises previstas é fundamental. Mas estudos grandes podem fornecer dados para um grande número de análises, principalmente para testar novas hipóteses. Passado esse período, os dados podem ser divulgados para uso por outros pesquisadores. A questão crucial é como fazer isso. Antes de abordar esse problema, toco em outro, que é o fato de que investigadores em países com estrutura de pesquisa menos desenvolvida ficam em uma situação de desvantagem em relação a grupos em países desenvolvidos. Essa desvantagem inclui a velocidade em produzir artigos, pois nos países mais desenvolvidos, as agências de fomento, geralmente incluem no financiamento dos projetos grandes somas de recursos para serem usadas na contratação de profissionais pós-graduados que irão atuar como auxiliares dos pesquisadores, e que acabam elaborando grande parte dos artigos, o que permite expressiva produção dos pesquisadores líderes. Essas possíveis desvantagens precisam ser levadas em conta ao se pensar em compartilhamento equitativo de dados.

Levando tudo o que foi exposto em conta, defendemos que o compartilhamento e publicização de dados produzidos a partir de estudos epidemiológicos em nosso país pode e deve ser uma meta. Para que isso ocorra de maneira adequada, precisamos de uma competência nacional para a devida preparação dos dados, apoiada por financiamento adequado por parte de nossas agências de fomento. Ou seja, o investigador precisa de incentivos adequados.^{1,7}

Não nos agrada a ideia de disponibilização total e aberta dos dados em um repositório. Nós temos muito a aprender com outros investigadores, mas eles, especialmente do exterior, também têm muito a aprender conosco. Desta forma, favorecemos a ideia de compartilhamento de dados a partir de solicitação e definição de objetivos da análise e o estabelecimento de uma colaboração, de forma semelhante ao que propõe Lopez.⁷ A participação de investigadores locais no processo de análise deve ficar a critério do grupo que fornece os dados. Isso deve levar em conta, principalmente, o interesse na análise proposta. Mas é claro que essa participação não deve se resumir a uma autoria sem real contribuição para o trabalho. O mundo desenvolvido realiza, há tempos, um processo de colonialismo científico, no qual pesquisadores financiados por agências de seus países realizam estudos e coletam dados em países pobres ou carentes de capacidade técnico-científica. Esses estudos servem para alavancar suas carreiras, gerar publicações e atrair alunos estrangeiros pagando altas taxas para as grandes instituições universitárias. O legado científico para o país alvo do estudo é bastante variável e discutível. Com sorte, o resultado da pesquisa irá beneficiar o país e sua população. Se passamos a distribuir nossos dados de forma totalmente aberta e irrestrita podemos estar contribuindo para este processo de forma ainda mais imediata.

Mas, como ressaltam Pearce & Smith,³ o compartilhamento equitativo, responsável e ético de dados não admite uma estratégia única. É necessário que particularidades do grupo que coletou os dados, do estudo, e dos próprios dados sejam levadas em conta para estabelecer critérios adequados. Defendemos o compartilhamento de dados, mas com critérios, e não compulsoriamente através de uma regra genérica.

Referências

1. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. Bull World Health Organ. 2010;88(6):462–6. doi: 10.2471/BLT.09.074393.
2. Bull S. Ensuring global equity in open research. London, UK: Wellcome Trust; 2016. doi: 10.6084/m9.figshare.4055181.

3. Pearce N, Smith AH. Data sharing: Not as simple as it seems. *Environ Heal A Glob Access Sci Source*. 2011;10(1):1–7. doi: 10.1186/1476-069X-10-107.
4. Langat P, Pisartchik D, Silva D, Bernard C, Olsen K, Smith M, et al. Is there a duty to share? Ethics of sharing research data in the context of public health emergencies. *Public Health Ethics*. 2011;4(1):4–11. doi: 10.1093/phe/phr005.
5. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for Responsible Sharing of Clinical Trial Data. Hamel MB, editor. *N Engl J Med*. 2013;369(17):1651–8. doi: 10.1056/NEJMhle1309073.
6. Domingo-ferrer J, Muralidhar K, Rufian-torrell G. Anonymization Methods for Taxonomic Microdata. In: Domingo-Ferrer J, Tinnirello I, editors. *Priv Stat Databases PSD 2012 Lect Notes Comput Sci*. Berlin: Springer; 2012. p. 90–102. doi: 10.1007/978-3-642-33627-0_8.
7. Lopez AD. Sharing data for public health: Where is the vision? *Bull World Health Organ*. 2010;88(6):467. doi: 10.2471/BLT.10.078956.