



Foto: Edson Ruiz/CIDACS. Reprodução

Sob liderança de Maurício Barreto, o CIDACS nasceu em 2016 para explorar os determinantes sociais e ambientais da saúde.

CIDACS: Transformando dados em ciência para a saúde pública

por Chris Bueno

A integração de dados para a produção de conhecimentos transformadores em saúde pública encontrou no Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), da Fiocruz Bahia, um exemplo pioneiro e inovador. Criado em 2016 sob a liderança de Maurício Barreto, médico e epidemiologista com uma carreira marcada pelo compromisso com a ciência de impacto social, o CIDACS se destaca mundialmente como uma das iniciativas mais avançadas na utilização de grandes bases de dados para promover equidade e bem-estar populacional. Maurício Barreto, PhD em Epidemiologia pela Universidade de Londres e membro da Academia Brasileira de Ciências (ABC), desempenhou um papel central na concepção do CIDACS, combinando sua vasta experiência acadêmica e científica com uma visão estratégica para explorar os determinantes sociais e ambientais da saúde. O centro, que abriga projetos de grande porte como a Coorte de 100 Milhões de Brasileiros, utiliza tecnologias de ponta e métodos interdisciplinares para investigar como políticas sociais e mudanças ambientais impactam a saúde e as desigualdades no Brasil. Reconhecido por seu protagonismo na ciência aberta, o CIDACS integra dados públicos em pesquisas que geram evidências científicas robustas, auxiliando na formulação de políticas públicas. Com contribuições significativas para o enfrentamento de crises, como a pandemia de Covid-19, e na defesa de legislações como a Lei Geral de Proteção de Dados (LGPD) voltada à saúde pública, o centro recebeu em 2023 o prêmio "The Information and Data Distribution Award", concedido pela Federação Mundial de Associações de Saúde Pública. Nesta entrevista, Maurício Barreto compartilha os desafios enfrentados para implantar o CIDACS, os avanços no uso de dados para pesquisas em saúde pública e o impacto transformador de iniciativas interdisciplinares e colaborativas no cenário científico brasileiro e internacional. Confira!

Ciência & Cultura – Conte um pouco sobre o projeto CIDACS – o que é, como nasceu, parcerias, principais resultados e desafios?

Maurício Barreto – O CIDACS tem cerca de 10 anos de história. Sou epidemiologista e sempre trabalhei com pesquisa em diversas áreas, utilizando inicialmente dados primários coletados diretamente em estudos. Mais tarde, começamos a usar dados dos sistemas nacionais de informação para avaliar políticas públicas. Por exemplo, participamos de um dos maiores ensaios clínicos do mundo, com mais de 300 mil pessoas, para avaliar os efeitos da revacinação com BCG. A partir desse trabalho, passamos a estudar o impacto de políticas públicas sobre a saúde usando dados agregados, como os dados municipais. No Brasil, com quase 6 mil municípios, é possível realizar análises robustas nesse formato. Investigamos, por exemplo, o programa Bolsa Família, uma política social que chamou nossa atenção devido à sua relação com os determinantes sociais de saúde, como pobreza e desigualdades. Em 2007, publicamos na *The Lancet* um artigo que mostrou o impacto positivo do Bolsa Família na saúde. O estudo teve grande repercussão e chamou a atenção da então ministra do Desenvolvimento Social, Tereza Campelo. Começamos a discutir novas possibilidades de estudo e conseguimos acesso a bases de dados identificadas, com as devidas proteções, transferidas para a Fiocruz, a Universidade Federal da Bahia (UFBA) e a Universidade de Brasília (UnB).

Com essa base de dados, estruturamos o CIDACS para armazenar, proteger e vincular informações de saúde e políticas sociais, como o Cadastro Único. Esse processo, conhecido como *data linkage*, é amplamente utilizado em países como Inglaterra, Canadá e Austrália, e começava a ser explorado no Brasil. Assim, nasceu a ideia de criar uma coorte com dados de 100 milhões de brasileiros, que hoje inclui cerca de 30 milhões de nascimentos. Vinculamos dados de saúde, mortalidade, nutrição, hospitalizações e políticas sociais, possibilitando análises detalhadas e inovadoras.

“Nossos estudos mostram claramente o impacto significativo da redução da pobreza na saúde.”

C&C – Qual foi a primeira informação relevante que observaram?

MB – Nossos estudos mostram claramente o impacto significativo da redução da pobreza na saúde. A melhora nas condições de saúde da população brasileira nas últimas décadas é notável, com redução da mortalidade infantil e da desnutrição, em parte devido a programas como o Bolsa Família, o Programa Cisternas e o Minha Casa, Minha Vida. Nosso foco principal foi o Bolsa Família. Demonstramos que, mesmo com transferências financeiras relativamente pequenas, o impacto na saúde e na qualidade de vida das famílias foi enorme.

Esses resultados estão bem documentados e representam uma das grandes contribuições do CIDACS. (Figura 1)

C&C – Em quais outros estudos o CIDACS tem se envolvido?

MB – O Cadastro Único (CadÚnico), que representa a população mais vulnerável do país, é uma base riquíssima para pesquisa. Ele nos permite investigar características demográficas e sociais específicas, como as condições de quilombolas e indígenas, além de medir desigualdades que ainda persistem, apesar das políticas sociais. Desenvolvemos estudos sobre doenças específicas, como infecções raras durante a gravidez, que exigem amostras muito grandes para serem analisadas adequadamente. Também exploramos as relações entre o Bolsa Família e a atenção primária à saúde, avaliando os efeitos de programas como a Estratégia Saúde da Família. Mais recentemente, nos aprofundamos nos efeitos das mudanças climáticas na saúde, com financiamento da *Wellcome Trust*, para integrar dados climáticos à nossa base nacional. O Brasil, com sua diversidade climática e biomas, oferece uma oportunidade única para investigar esses efeitos. Por exemplo, queremos entender o impacto de políticas como o Programa Cisternas no enfrentamento da seca. Além disso, estamos desenvolvendo sistemas para detecção precoce de epidemias, utilizando nossa expertise em integração de dados. Isso é fundamental, considerando a possibilidade concreta de novas epidemias globais.



Foto: Matheus Britto/Prefeitura Municipal do Jaboatão dos Guararapes. Reprodução.

Figura 1. Programas sociais, como o Bolsa Família, transformaram a saúde no Brasil, reduzindo mortalidade infantil e desnutrição.

C&C – Sua pesquisa e trabalho de décadas em saúde pública são fortemente baseados na análise de dados de saúde, tanto abertos quanto disponibilizados via projetos com restrições de acesso. Como é tratada a questão da privacidade e proteção desses dados?

MB – Desde o início, o CIDACS teve como prioridade a privacidade e a proteção dos dados, o que resultou em um grande investimento em segurança. Contamos com grupos especializados em governança de dados, como o liderado pela doutora Betânia Almeida, que é referência internacional na área. Além disso, temos uma equipe técnica dedicada à curadoria e à segurança de dados. Por exemplo, desenvolvemos a primeira curadoria de dados deste porte (100 milhões de brasileiros), para saúde pública, no Brasil, voltada para pesquisa, com controle rigoroso de movimentação dos dados. Qualquer dado

identificado é tratado em ambientes isolados e desconectados da internet, eliminando riscos de invasões ou vazamentos. Quando um pesquisador apresenta uma questão científica, ele deve submeter um plano de dados que passa por análise da curadoria e aprovação ética. Só após essa etapa, os dados desidentificados podem ser acessados em áreas seguras, via VPN, sem possibilidade de *download*, nem mesmo de dados desidentificados. Pesquisadores em diferentes estados e países – como Minas Gerais, Rio de Janeiro, Londres e Glasgow – acessam os dados usando sistemas altamente seguros, mas os dados permanecem no CIDACS. Trabalhamos com práticas de segurança internacionalmente reconhecidas e, até hoje, nunca enfrentamos incidentes de vazamento. O acesso é controlado: os pesquisadores podem gerar tabelas e gráficos, mas não têm contato direto com os dados individuais. Além disso,

o CIDACS atua como modelo na área de governança e curadoria de dados, o que nos enche de orgulho. Nossa abordagem exemplifica como é possível manter altos padrões de segurança e ainda produzir ciência relevante e impactante. (Figura 2)

C&C – Como o CIDACS e outros grandes projetos de dados de saúde firmam parcerias para uso controlado dessas informações?

MB – Estabelecemos parcerias de diferentes tipos. Algumas são voltadas para a transferência de tecnologia e conhecimento; outras, para responder conjuntamente a questões científicas. Nesses casos, colaboramos com pesquisadores para captar recursos e fortalecer nossa infraestrutura. Importante ressaltar que os dados nunca saem do CIDACS – os pesquisadores vêm até o sistema para realizar análises. Temos, por exemplo, uma parceria interessante com a África do Sul para comparar estruturas de dados e explorar como as perguntas e respostas científicas variam conforme o contexto. Isso reflete o fato de que dados não são neutros; eles são construídos em contextos específicos. Outra linha de parceria envolve o desenvolvimento de modelos comuns de dados, fundamentais para análises federadas. Nesse modelo, testamos hipóteses em diferentes contextos sem transferir dados, integrando os resultados por meio de meta-análises. Utilizamos essa abordagem em estudos sobre vacinas da COVID-19 com colegas da Escócia e do Brasil. Essas iniciativas demonstram a

qualidade dos dados brasileiros e sua relevância científica, algo que é constantemente validado por publicações em revistas internacionais de alto impacto. Nosso trabalho comprova que, mesmo enfrentando limitações tecnológicas, é possível criar soluções inovadoras e colaborar ao nível global, especialmente entre países do Sul Global.

C&C – O grande volume de dados com que vocês trabalham pode dificultar diagnósticos ou análises? Ferramentas de inteligência artificial (IA) são utilizadas para auxiliar nesse processo?

MB – Embora a inteligência artificial tenha grande potencial, ela ainda não é central nas nossas operações. Alguns pesquisadores utilizam ferramentas de IA, mas nosso foco principal é responder a questões científicas, para as quais existem métodos e algoritmos já consolidados. Trabalhamos com ferramentas como Python e R para análises estatísticas em larga escala e manipulação de grandes volumes de dados. Um dos desafios que enfrentamos é o *linkage* de dados. No Brasil, a utilização crescente do CPF como identificador universal tem facilitado a conexão entre diferentes bases de dados. Porém, muitos sistemas ainda utilizam informações como nome e nome da mãe, o que exige o desenvolvimento de *softwares* específicos para vincular os dados. Criamos nosso próprio *software* para isso, e estamos avaliando métodos que integrem IA para aprimorar essas soluções. Porém, é importante destacar que *big data* e inteligência artificial



Foto: Ascom CIDACS. Reprodução.

Figura 2. Referência global, o CIDACS usa grandes dados e tecnologia para promover saúde, equidade e reduzir desigualdades no Brasil.

não são sinônimos. Vejo uma confusão crescente no Brasil, onde se discute IA sem considerar a base de dados. Isso pode gerar distorções significativas. Dados de qualidade e bem estruturados são o alicerce para qualquer aplicação de IA. Ainda há um longo caminho para explorar o potencial da IA em questões científicas, mas precisamos avançar com cuidado e planejamento.

C&C – A Ciência Aberta é baseada, entre outros princípios, no compartilhamento livre de dados por meio de sua disponibilização em repositórios públicos. No caso de dados envolvendo questões legais ou éticas (como os associados a pessoas), há o compromisso de serem “tão abertos quanto possível, tão fechados quanto

necessário”. Como isso é praticado no CIDACS e em outros projetos dos quais você participa?

MB – Um ponto importante é a divulgação dos achados científicos. Todas as nossas publicações são em acesso aberto, e essa é uma política nossa. Contamos com um sistema de colaboração intensa, em que pesquisadores podem acessar os dados para responder a questões acadêmicas. No entanto, tudo isso está sujeito a aprovações éticas e à curadoria dos dados, pois não há possibilidade de abrir completamente dados sensíveis, como informações de saúde, para uso irrestrito. Esse modelo seria incompatível com a ética, a Lei Geral de Proteção de Dados (LGPD) e as normas do sistema ético nacional, como as da CONEP. Dados

sensíveis estão regulados por leis e sistemas éticos, tanto no Brasil quanto em outros países, controlando sua utilização. A abertura de dados, nesse contexto, é diferente: o acesso pode ser permitido para responder a uma questão relevante, mas sempre dentro de um sistema fechado e sob supervisão. Por outro lado, os resultados das pesquisas são amplamente divulgados em acesso aberto. As bases de dados são codificadas e podem ser testadas dentro desse sistema, mas não há possibilidade de torná-las completamente públicas. Adicionalmente, temos um núcleo dedicado à disseminação de informações, com ampla participação de diferentes representações sociais. Muitos dos nossos projetos incluem comitês que revisam e discutem as propostas. Também realizamos estudos com gestores e membros da sociedade para compreender suas perspectivas sobre a utilização de dados públicos. No entanto, é importante destacar que os dados não pertencem a nós, pesquisadores; eles pertencem à sociedade brasileira e são regulados por leis e sistemas éticos. O nível de abertura

“Nosso trabalho comprova que, mesmo enfrentando limitações tecnológicas, é possível criar soluções inovadoras e colaborar em nível global, especialmente entre países do Sul Global.”

é decidido por esses sistemas, não pelos pesquisadores. O que podemos controlar é a divulgação das publicações e a criação de infraestrutura que facilite o acesso a grupos interessados. Esse processo não é simples. Envolve custos e sistemas operacionais complexos. É por isso que os dados permanecem sob proteção rigorosa, mesmo quando desidentificados, e ainda não consideramos seguro transferi-los para outros ambientes. Priorizamos a privacidade acima do conceito de acesso aberto. Estamos também desenvolvendo iniciativas

para abrir dados rotineiros que não apresentam questões éticas, como dados climáticos e informações já disponíveis pelo Ministério da Saúde. Esses dados serão documentados, curados e disponibilizados em formatos acessíveis em um portal para uso por pesquisadores, gestores ou qualquer pessoa interessada. Outra inovação em que trabalhamos é a criação de coortes sintéticas, como a população brasileira sintética desenvolvida por um pesquisador do CIDACS. Essa população simula as características da população brasileira, mas não representa indivíduos reais. Com 190 milhões de pessoas fictícias, ela pode ser usada para análises preliminares, projeções e modelos sintéticos em diversas áreas científicas. É uma forma de compensar as limitações do acesso a dados reais e sensíveis.

Chris Bueno é jornalista, escritora, divulgadora de ciências, editora-executiva da revista Ciência & Cultura, e mãe apaixonada por escrever (especialmente sobre ciência).