

USO DE BASES DE DADOS EM SAÚDE E A PRIVACIDADE E PROTEÇÃO DE DADOS PESSOAIS

Use of health databases and privacy and personal data protection

Néa Miwa Kashiwagi¹

Zillá Oliva Roma²

SUMÁRIO

1. Introdução; 2. A proteção de dados no Brasil; 3. Os dados de saúde; 4. A proteção de dados em pesquisas de saúde; 5. Anonimização e pseudoanonimização de base de dados; 6. Conclusões; Referências Bibliográficas.

RESUMO

Bases de dados secundárias têm sido fontes de informação importantes para o desenvolvimento de pesquisas em saúde, cujos resultados irão contribuir para o aprimoramento de protocolos assistenciais e políticas públicas. A Lei Geral de Proteção de Dados (LGPD), além de consolidar conceitos já presentes em normas técnicas e códigos de ética existentes no campo da pesquisa, tais como o sigilo de dados e a autodeterminação informativa, trouxe um novo olhar sobre as novas demandas tecnológicas relacionadas ao uso massivo de informações. Sendo assim, a utilização de técnicas de anonimização e pseudoanonimização para o tratamento de dados pessoais deve ser aplicada de forma que busque o equilíbrio entre garantir o sigilo dos dados e não comprometer a qualidade da informação. Para isso, é necessário o desenvolvimento de algoritmos e metodologias de controle de dados científicos. Algumas iniciativas têm sido desenvolvidas, como os Centros de Dados orientados para pesquisa, que fornecem infraestrutura e profissionais especializados em gerenciar diferentes bases de dados, e os movimentos de compartilhamento de dados abertos que, em consonância com as medidas de proteção de dados pessoais, disponibilizam dados para incentivar o desenvolvimento de pesquisas em saúde.

Palavras-chave: Proteção de Dados. LGPD. Pesquisa em Saúde. Base de Dados. Saúde Pública.

ABSTRACT

Secondary databases are important sources of information for the development of health research, whose results will contribute to the improvement of care protocols and public policies. The Brazilian General Data Protection Law (LGPD), in addition to consolidating concepts already present in existing technical standards and ethical codes in the field of research, such as data confidentiality and informational self-determination, has brought a new perspective on the technological demands of massive information use. Therefore, the use of anonymization and pseudo-anonymization techniques for processing personal data must aim to strike a balance between ensuring data confidentiality without compromising information quality. This requires the development of algorithms and methodologies for managing

¹ Enfermeira do Centro de Vigilância Sanitária da Secretaria de Estado da Saúde de São Paulo e Analista em Epidemiologia do Hospital Israelita Albert Einstein. Mestre em Medicina Preventiva pela Universidade de São Paulo. Especialista em Administração Hospitalar pela Universidade de São Paulo.

² Procuradora do Estado de São Paulo. Doutora e Mestra em Direito Processual Civil pela Universidade de São Paulo (USP). Autora do livro "Tutela jurisdicional e direito à saúde: análise crítica da farmacização do Judiciário" e de artigos jurídicos. Membro da Comissão Especial de Direito Constitucional da OAB/SP. Professora em cursos diversos. Palestrante.

scientific data. Some initiatives have been developed, such as Research-Oriented Data Centers, which provide infrastructure and specialized professionals to manage different databases, and Open Data Sharing movements, which make data available in compliance with personal data protection measures to support the development of health research.

Keywords: Data Protection. General Data Protection Regulation. Health Services Research. Database. Public Health.

1. INTRODUÇÃO

O artigo de Warren e Brandeis intitulado “The right to privacy”³ foi o marco inicial do conceito de privacidade, cuja expressão se pautou no individualismo e no direito de ser deixado só. Por ter características patrimonialistas, restringia-se às classes tipicamente burguesas. No entanto, essa visão, corroborada pelo crescimento do fluxo de informações digitais, constituiu-se como pré-condição fundamental para o exercício de diversas outras liberdades fundamentais⁴.

A partir do conceito inicial de privacidade associado ao isolamento, refúgio ou segredo, a noção de privacidade recebe influências sociais, como a relação do indivíduo com o Estado, as reivindicações de classes trabalhadoras por direitos e, juntamente com o aumento da capacidade técnica de processamento de informações, que moldaram e ampliaram o uso de informações pessoais, amparadas pela necessidade de eficiência e controle⁵.

O Estado foi o primeiro a coletar, tratar e utilizar dados pessoais sob a justificativa de uma Administração Pública mais eficiente. No entanto, quanto mais se conhece os cidadãos, maior o poder estatal sobre eles, sendo esse forte controle social característico de regimes totalitários⁶.

A tecnologia trouxe, ao longo dos anos, maior rapidez na coleta e no tratamento das informações. Seu avanço permitiu que qualquer instrumento, desde um celular até uma simples torradeira, seja capaz de captar, coletar, armazenar, processar e analisar informações dos indivíduos, empresas e governos durante as atividades corriqueiras. Além disso, esse progresso contribui para a formatação de produtos personalizados, os quais vêm sendo gradativamente substituídos por mercados ricos em dados.

³ WARREN, Samuel; BRANDEIS, Louis. The right to privacy. *Harvard Law Review*, Cambridge, v. IV, n. 15, dez. 1890. Disponível em: https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html. Acesso em: 3 abr. 2026.

⁴ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. 3. ed. São Paulo: Thomson Reuters Brasil, 2021. p. 30-47.

⁵ Ibid.

⁶ Ibid.

Consequentemente, muitas informações não são coletadas por necessidade, mas pela facilidade em sua obtenção e “[...] independentemente do paradigma tecnológico, o controle da informação foi sempre um elemento essencial na definição de poderes dentro de uma sociedade [...]”^{7 8}.

O campo da saúde vem acompanhando essa transformação com dados que passam a ser digitais de forma crescente e com o desenvolvimento de tecnologias, como o *Big Data* (grandes dados ou megadados), incorporando técnicas de análise sofisticadas que possibilitam o desenvolvimento de novas formas de produção de conhecimento em escala.

Contudo, ao mesmo tempo em que grandes bancos de dados podem fornecer insumos para o desenvolvimento de conhecimentos científicos que beneficiam a sociedade e o desenvolvimento de políticas públicas, emergem riscos de reidentificação⁹ e possíveis ataques cibernéticos que podem comprometer tanto a proteção dos participantes da pesquisa, quanto a autonomia, a privacidade e a confidencialidade de suas informações.

O presente estudo tem como objetivo compreender a relação entre dados utilizados em pesquisas de saúde e a privacidade e proteção de dados pessoais, buscando-se explorar os possíveis caminhos de convergência para garantir tanto o direito à saúde, quanto o direito à privacidade da população na utilização de novas tecnologias de uso massivo da informação.

2. A PROTEÇÃO DE DADOS NO BRASIL

As primeiras discussões sobre a necessidade de regulamentos sobre privacidade e proteção de dados no mundo remontam a meados da década de 1960,

⁷ Ibid.

⁸ ABRAMOVAY, Ricardo; ZANATTA, Rafael Augusto Ferreira. Dados Pessoais Abertos: Pilares dos Novos Mercados Digitais? *RDU*, Porto Alegre, v. 16, n. 90, p. 155-178, 2019.

⁹ A reidentificação consiste no processo de reassociar dados anonimizados ou pseudoanonimizados a um indivíduo específico, representando um sério risco à privacidade e à conformidade com a LGPD. O uso de inteligência artificial (IA) e o cruzamento de bases de dados (quase identificadores) tornam a reidentificação possível, exigindo técnicas robustas de proteção, como a anonimização eficaz e a privacidade diferencial. “A proliferação de informações publicamente disponíveis online, combinada com hardware de computador cada vez mais poderoso, tornou possível a reidentificação de dados ‘anonimizados’. Isso significa que dados apagados agora podem ser rastreados até o usuário individual a quem se referem. A reidentificação de dados apagados é comumente realizada combinando dois ou mais conjuntos de dados para encontrar o mesmo usuário em ambos. Essa informação combinada frequentemente revela informações de identificação direta sobre um indivíduo. A reidentificação de dados anonimizados tem sérias implicações para a privacidade e para as políticas públicas, à medida que reguladores, empresas e consumidores lutam para definir privacidade na era moderna do registro permanente”. LUBARSKY, Boris. Reidentificação de dados “anonimizados”. *Georgetown Law Technology Review*, 2017. Disponível em: https://georgetownlawtechreview-org.translate.goog/re-identification-of-anonymized-data/GLTR-04-2017?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc. Acesso em: 3 abr. 2026.

tendo sido a Alemanha o primeiro país a elaborar um marco regulatório sobre o tema, em 1970¹⁰.

Nos últimos 20 anos, vários países produziram normas, totalizando 142 países com legislação específica sobre privacidade e proteção de dados¹¹. No Brasil, a privacidade e a proteção de dados pessoais estão contempladas na Constituição Federal sob as garantias de: liberdade de expressão (art. 5º, inc. IX; art. 220), direito à informação (art. 5º, inc. XIV; art. 220), salvaguarda da vida privada e da intimidade (art. 5º, inc. X) e do domicílio (art. 5º, inc. XI), bem como o sigilo de sua correspondência, das comunicação telegráficas e telefônicas e de seus dados (art. 5º, inc. XII), e a concessão de *habeas-data* (art. 5º, inc. LXII).

Outras legislações, como o Código Civil (art. 23)¹² e o Código de Defesa do Consumidor (CDC)¹³, trazem algumas garantias de proteção à vida privada e às informações cadastrais. No entanto, mostravam-se insuficientes, principalmente diante das novas circunstâncias decorrentes da evolução tecnológica, tais como o uso abusivo de informações pessoais armazenadas em bancos de dados, que poderiam ser utilizados para fins diversos daqueles que motivaram sua coleta¹⁴.

A Emenda Constitucional nº 115, de 10 de fevereiro de 2022, a fim de assegurar “a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais [...]”¹⁵, incluiu dois incisos no art. 5º da Constituição: o LXXIX, que garante o direito à proteção de dados pessoais, inclusive em meios digitais; e o XXVI que determina a criação de ações de fiscalização dos tratamentos de dados pessoais; ao art. 22, acresceu a responsabilidade da União diante da proteção e do tratamento de dados pessoais, conforme o inciso XXX.

Em 14 de agosto, de 2018, com a promulgação da Lei Geral de Proteção de Dados (LGPD)¹⁶, estruturou-se a base normativa dedicada à disciplina da proteção

¹⁰ DONEDA, Danilo. A LGPD como elemento estruturante do modelo brasileiro de proteção de dados. In: **Lei geral de proteção de dados (Lei nº 13.709/2018): a caminho da efetividade: contribuições para a implementação da LGPD**. São Paulo: Thomson Reuters; Revista dos Tribunais, 2020, p. 243-255.

¹¹ Ibid.

¹² BRASIL. Lei nº 10.406, de 10 de janeiro de 2002. Institui o Código Civil. **Diário Oficial da União**, Brasília, DF, 10 jan. 2002.

¹³ BRASIL. Lei nº 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências. **Diário Oficial da União**, Brasília, DF, 1990.

¹⁴ DONEDA, 2021, *op. cit.*, p. 267-368.

¹⁵ BRASIL. Emenda Constitucional nº 115, de 10 de fevereiro de 2022. Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais. **Diário Oficial da União**, Brasília, 2022.

¹⁶ BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 17 jun. 2024.

de dados, na qual se estabeleceram os princípios de proteção aplicáveis a diferentes situações de tratamento de dados pessoais, garantindo-se a proteção dos direitos fundamentais de liberdade e personalidade¹⁷.

A LGPD, além de consolidar elementos já presentes na ordem jurídica, incluiu conceitos novos e próprios de proteção de dados, seus princípios, os direitos do titular e as regras de tratamento que consideram o risco dessa atividade. A lei traz um instrumental jurídico próprio, sendo aplicável em diferentes atividades econômicas e instituições, sejam elas públicas ou privadas, em diferentes cenários e situações, incorporando-se à sociedade os parâmetros utilizados globalmente sobre regulação e proteção de dados¹⁸.

3. OS DADOS DE SAÚDE

A utilização de bases de dados de saúde como instrumento de apoio na identificação de epidemias que assolam a sociedade não é recente. Estudos clássicos, como o surto de cólera em Londres, em 1854, em que o médico John Snow estudou sua transmissão a partir do mapeamento das casas com enfermos e da identificação de determinado poço de água como fonte de contaminação; bem como Florence Nightingale, que demonstrou em gráficos que a principal causa de morte entre os soldados da Guerra da Criméia foi infecção por feridas de guerra não tratadas, são exemplos de que a aplicação de dados por métodos estatísticos e epidemiológicos pode ser essencial para se compreender e definir estratégias de combate às doenças^{19 20}.

A Epidemiologia, enquanto campo científico que estuda a dinâmica da saúde e do adoecimento das populações, a partir dos anos 1960, passou por uma transformação histórica de suas práticas com o desenvolvimento de modelos matemáticos e técnicas analíticas complexas, graças à introdução da computação eletrônica, que permitiu a análise de grandes bases de dados e a sistematização do conhecimento epidemiológico²¹.

No Brasil, os primeiros passos para a construção de bases de dados de saúde começaram em 1977, com a padronização dos registros de óbito e a criação de um sistema informatizado para a coleta de dados dos atestados de óbito e a formatação dos dados de produção hospitalar pelo extinto Instituto Nacional de Assistência Médica da Previdência Social (Inamps). Ao longo do tempo, essas iniciativas foram

¹⁷ DONEDA, 2020, *op. cit.*

¹⁸ Ibid.

¹⁹ SALSBURG, David. **Uma senhora toma chá...**: Como a estatística revolucionou a ciência no século XX. São Paulo: Zahar, 2009.

²⁰ SNOW, John. **Sobre a maneira de transmissão do cólera**. São Paulo: Hucitec, 1999.

²¹ ALMEIDA FILHO, Naomar de. Bases históricas da Epidemiologia. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 2, n. 3, p. 304-311, set. 1986.

aprimoradas e se tornaram o Sistema de Informações de Mortalidade (SIM) e o Sistema de Internações Hospitalares (SIH-SUS)²².

Em 1988, a criação do Sistema Único de Saúde (SUS) trouxe a garantia constitucional do direito à saúde que se materializa em forma de políticas públicas voltadas para a redução de riscos e agravos à saúde, bem como para o acesso às ações e serviços públicos de saúde, prioritariamente preventivos, mas sem prejuízo das atividades assistenciais. Logo, o Sistema Nacional de Vigilância em Saúde (SNVS) tem um papel importante no desenvolvimento da coleta de dados, na consolidação e na análise de informações que identifiquem riscos e agravos à saúde e que sirvam como subsídios para o planejamento de medidas de promoção da saúde e redução de morbidades²³.

A Política Nacional em Vigilância em Saúde (PNVS), como política pública de Estado e função essencial do SUS, orienta o modelo de atenção, no qual as ações de vigilância em saúde são definidas em todos os níveis de atenção do SUS e a Vigilância em Saúde é descrita como “o processo contínuo e sistemático de coleta, consolidação, análise de dados e disseminação de informações sobre eventos relacionados à saúde”, que visam ao “planejamento e à implementação de medidas de saúde pública, incluindo a regulação, intervenção e atuação em condicionantes e determinantes da saúde, para a proteção e promoção da saúde da população, prevenção e controle de riscos, agravos e doenças”, conforme § 1º do art. 2º da Resolução n. 588/2018²⁴.

Além do SIM e do SIH, podemos citar o Sistema Nacional de Notificação Compulsória (SINAN), que consiste na comunicação obrigatória das doenças de notificação compulsória à autoridade de saúde, realizada por profissionais da área (médicos, enfermeiros, fisioterapeutas, entre outros) ou pelo responsável pelo estabelecimento de saúde, seja público ou privado, sobre a ocorrência de suspeita ou confirmação de doença, agravo ou evento de saúde pública.

Esse sistema de alerta é primordial para a identificação de emergências de saúde pública que representam risco imediato à saúde e que demandam a implantação de estratégias oportunas para a contenção da doença. Um exemplo de monitorização de doenças de grande impacto na saúde pública foi a identificação dos primeiros casos de covid-19 no Brasil, que motivaram as primeiras medidas de isolamento social como forma de contenção da doença²⁵.

²² LIMA, Claudia Risso de Araujo *et al.* Departamento de Informática do SUS – DATASUS: a experiência de disseminação de informações em saúde. In: BRASIL (org.). **A experiência brasileira em sistemas de informação em saúde**. Brasília: Editora do Ministério da Saúde, 2009. v. 1, p. 109-128.

²³ *Ibid.*

²⁴ BRASIL. Conselho Nacional de Saúde. **Resolução nº 588, de 12 de julho de 2018**. Institui a Política Nacional de Vigilância em Saúde (PNVS). Brasília, DF: Conselho Nacional de Saúde, 2018. Disponível em: <https://conselho.saude.gov.br/resolucoes/2018/Reso588.pdf>. Acesso em 3 abr. 2026.

²⁵ BRASIL. Ministério da Saúde. **Portaria nº 188, de 3 de fevereiro de 2020**. Declara Emergência em Saúde Pública de importância Nacional (ESPIN) em decorrência da Infecção Humana pelo novo Coronavírus (2019-nCoV). Brasília, DF: Ministério da Saúde, 2020.

O Sistema de Informações de Nascidos Vivos (SINASC), implantado pelo Ministério da Saúde, em 1990, tem como objetivo garantir o registro sistemático de informações de nascidos vivos no país. A Declaração de Nascido Vivo (DN) é o documento padrão de preenchimento obrigatório em todo o território nacional, cuja emissão e distribuição são de responsabilidade do Ministério da Saúde, enquanto as Secretarias Estaduais e Municipais de Saúde realizam a distribuição das DN nas maternidades e controlam o fluxo de informação das DN entre os estabelecimentos de saúde, cartórios de registro civil e profissionais de saúde autônomos que realizam partos domiciliares²⁶.

Em três vias pré-numeradas sequencialmente, os estabelecimentos de saúde realizam o preenchimento da via impressa com dados da mãe (idade, raça e atividade econômica), pré-natal (número de consultas de pré-natal), parto (tipo de parto, idade gestacional) e recém-nascido (peso, altura, Apgar ao nascimento), e inserem as informações em uma base de dados digital. Tais informações são utilizadas pelos gestores municipais e estaduais como subsídios para o planejamento de políticas públicas, além de construírem uma base de dados amplamente utilizada em pesquisas científicas²⁷.

Em um estudo de revisão sobre o SINASC²⁸, o acesso amplo de suas bases de dados possibilitou o desenvolvimento de diferentes publicações científicas por instituições renomadas de pesquisa e por serviços de saúde assistenciais, que desenvolvem análises temporais, estudos de prevalência, diagnósticos de saúde e desenvolvimento de indicadores de qualidade. Como exemplo, podemos citar o acompanhamento da taxa de mortalidade neonatal, que, do total de nascimentos no período, apresenta a proporção de crianças que morrem na primeira semana de vida, sendo que altas taxas podem sinalizar falhas na assistência pré-natal, parto e risco à saúde do nascituro.

Além das bases de dados voltadas à vigilância em saúde, o prontuário médico é uma importante fonte de informação da saúde dos indivíduos. Trata-se de um documento utilizado por estabelecimentos assistenciais de saúde, no qual são registrados dados do paciente: demográficos, queixas de saúde, história familiar e pessoal, estilo de vida, procedimentos de saúde, progressos, histórico de doenças, uso de medicamentos, resultados de exames e procedimentos, entre outros. Essas informações possibilitam que a equipe assistencial registre, monitore e avalie a evolução do paciente frente a uma doença ou risco de saúde, e defina condutas assistenciais que levem à recuperação da saúde do paciente.

²⁶ BRASIL. Ministério da Saúde. **Sistemas de informação em saúde**. 2024. Disponível em: <https://www.gov.br/aids/pt-br/svsa-na-gestao-municipal/agrivos-e-doencas-cronicas-nao-transmissiveis/links-de-interesse-sobre-agrivos-e-doencas-cronicas-nao-transmissiveis/sistemas-de-informacao-em-saude>. Acesso em: 17 jun. 2024.

²⁷ *Ibid.*

²⁸ PAIVA, Natália Santana, *et al.* Sistema de informações sobre nascidos vivos: um estudo de revisão. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 16, supl. 1, p. 1211-1220, 2011.

Para garantir a qualidade do registro dos dados, o Conselho Federal de Medicina (CFM) definiu os itens obrigatórios na composição de um prontuário²⁹. No entanto, cada serviço de saúde tem autonomia para elaborar seus próprios formulários, a forma de registro dos dados e o sistema de informação que irá adotar quando o prontuário é digital. Ainda que haja uma heterogeneidade de sistemas no mercado e que, em geral, não se comunicam, os sistemas de inteligência artificial possibilitarão, em breve, a associação de dados hospitalares de diferentes serviços ou bases de dados diversas de forma precisa e rápida.

Na área de saúde, as pesquisas científicas possuem diferentes metodologias. Em relação ao tipo de coleta de dados, duas formas são apresentadas: pesquisas com base de dados primários, quando estes são coletados com a finalidade específica da pesquisa, isto é, a metodologia de coleta, tendo os formulários elaborados somente para a pesquisa em desenvolvimento; já a pesquisa com base de dados secundárias utiliza fontes já existentes, construídas para outra finalidade (dados demográficos, epidemiológicos, clínicos, diagnósticos e terapêuticos), e que podem vir a ser utilizadas para o desenvolvimento de uma pesquisa científica.

Pela própria construção do raciocínio clínico que provém da associação de diversos fatores de risco, que podem provocar e definir a evolução de uma doença ou situação de saúde, a coleta de dados na área da saúde dispõe, além do cotidiano da assistência à saúde, de dados que podem consolidar indicadores e fornecer informações de apoio à gestão de saúde e elaboração de fontes de dados para o desenvolvimento de pesquisas científicas. A utilização de bases de dados secundárias, como fonte de informação para o desenvolvimento de pesquisas científicas, tem sido uma opção mais acessível e ágil em relação às pesquisas com base de dados primários, que, pela própria configuração, demandam maior volume de recursos financeiros e recursos humanos treinados e específicos.

Logo, verifica-se que há a necessidade de se criar uma agenda de tratamento de dados sem violar a privacidade e a proteção de dados. Da mesma forma que a circulação da informação pode ser benéfica por possibilitar o desenvolvimento de pesquisas e políticas públicas que contribuirão com a saúde, podendo oferecer mais oportunidades aos cidadãos; por outro lado, ela pode trazer aspectos negativos, como

²⁹ A Resolução CFM 1.638/2002 do Conselho Federal de Medicina, no art. 5º, define os itens obrigatórios no prontuário médico, seja eletrônico ou de papel: “[...] a. Identificação do paciente – nome completo, data de nascimento (dia, mês e ano com quatro dígitos), sexo, nome da mãe, naturalidade (indicando o município e o estado de nascimento), endereço completo (nome da via pública, número, complemento, bairro/distrito, município, estado e CEP); b. Anamnese, exame físico, exames complementares solicitados e seus respectivos resultados, hipóteses diagnósticas, diagnóstico definitivo e tratamento efetuado; c. Evolução diária do paciente, com data e hora, discriminação de todos os procedimentos aos quais o mesmo foi submetido e identificação dos profissionais que os realizaram, assinados eletronicamente quando elaborados e/ou armazenados em meio eletrônico [...]”. CONSELHO FEDERAL DE MEDICINA. Resolução nº 1.638/2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. **Diário Oficial da União**, Brasília, 2002.

incidentes de segurança e a consequente exposição de dados pessoais. Na saúde, em especial, a divulgação de dados sensíveis pode causar constrangimentos, discriminações e outras consequências danosas aos seus titulares.

Algumas frentes de promoção ao compartilhamento de dados em prol da ciência têm se empenhado em estabelecer diretrizes que busquem o equilíbrio entre o uso de bases de dados secundárias e a proteção dos dados.

O Movimento da Ciência Aberta, por exemplo, é uma iniciativa de financiadores de pesquisas científicas, governamentais e não-governamentais de vários países, que buscam promover o acesso aberto a publicações e dados científicos gerados em pesquisas, em especial as que são geradas com recursos públicos, com o intuito de disseminar conhecimentos que beneficiem a sociedade como parte integrante de políticas públicas. Para isso, essas frentes elaboram princípios e recomendações para garantir a convergência e a padronização do tratamento e do compartilhamento dos dados, de acordo com princípios éticos e regulatórios aceitos globalmente³⁰.

4. A PROTEÇÃO DOS DADOS EM PESQUISAS DE SAÚDE

A privacidade, o uso e o tratamento de dados pessoais sempre foram motivo de preocupação no campo da pesquisa em saúde. O Código de Nuremberg (1947), como resposta aos cruéis experimentos com seres humanos ocorridos durante o Holocausto, trouxe os princípios éticos que permeiam a pesquisa envolvendo seres humanos, dentre eles, o consentimento voluntário. Logo, ao participante deve ser assegurado o direito ao conhecimento e à compreensão da pesquisa e o direito à livre escolha, pelo qual se garante autonomia em suas decisões. Em 1975, a privacidade foi inserida na Declaração de Helsinque, na qual se determinou a responsabilidade dos pesquisadores de adotarem medidas para garantir a privacidade, a integridade física e mental, e a personalidade do indivíduo. O conceito de confidencialidade das informações foi incorporado a essa declaração em 2000³¹.

No Brasil, a privacidade dos sujeitos da pesquisa e o sigilo das informações foram definidos na Resolução CNS nº 466/2012³² do Conselho Nacional de Saúde. Entre os fundamentos éticos, a confidencialidade, a privacidade e a proteção da imagem e da pessoa são citados no item III.2.i, o qual prevê que o pesquisador

³⁰ ALMEIDA, Betânia de Araújo; PIMENTA, Denise Moraes; BARRETO, Mauricio Lima. Compartilhamento e abertura de dados na Ciência Aberta e a pesquisa em saúde pública na Lei Geral de Proteção de Dados. **Bahia Análise & Dados**, Salvador, v. 31, n. 2, p. 78-92, 2021.

³¹ DALLARI, Analluza Bolivar. Enquadramento do contrato de pesquisa clínica à luz das normas de aplicação imediata brasileiras. In: DALLARI, Analluza Bolivar (org.). **Contrato de Pesquisa Clínica: aspectos práticos e jurídicos**. São Paulo: Revista dos Tribunais, 2019, p. 163-2010.

³² BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução nº 466, de 12 de dezembro de 2012**. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília, DF: Ministério da Saúde, 2012.

é responsável por “[...] manter os dados da pesquisa em arquivo, físico ou digital, sob sua guarda e responsabilidade, por um período de cinco anos após o término da pesquisa”, conforme o item XI.2.f.

O Código de Ética Médica³³ traz definições importantes sobre o prontuário médico, não podendo o profissional negar ao paciente e seus familiares o acesso aos prontuários, ou deixar de informá-los sobre seu estado de saúde e riscos envolvidos. O médico deve manter sigilo das informações resultantes do desempenho de suas funções e as informações do prontuário devem ser guardadas de forma sigilosa. No âmbito da pesquisa, o mesmo Código cita a responsabilidade desse profissional ao participar de pesquisas que envolvem seres humanos ou qualquer outro animal, devendo seguir as normas éticas brasileiras e proteger a vulnerabilidade dos sujeitos da pesquisa.

Muitos aspectos contidos na LGPD, como o respeito à privacidade, o sigilo dos dados e a autodeterminação informativa, já faziam parte da rotina das pesquisas científicas. O Termo de Consentimento Livre e Esclarecido, por exemplo, sempre foi instrumento obrigatório aplicado em pesquisas com dados primários, garantindo autonomia, acesso às informações e proteção da privacidade dos participantes da pesquisa. Sendo assim, observa-se uma convergência entre as normativas do CEP/CONEP com a LGPD. No entanto, são necessários esforços institucionais para a adequação de situações novas como o *Big Data*³⁴.

A pesquisa com bases de dados secundárias, dada a categorização de suas variáveis e a apresentação dos resultados de forma agregada, não faria referência a dados individuais de seus participantes. No entanto, tais pesquisas podem trazer um risco de discriminação não direcionado a uma pessoa em particular, mas a um grupo de pessoas com características em comum (status social, econômico, etnia, gênero) e, também, em algumas situações em que a identificação pessoal se faz necessária para agregar dados de diferentes fontes de informação, sob risco de exposição de dados sensíveis. Esses riscos podem ser catalisados, ainda, por algoritmos de reidentificação e tais situações devem ser avaliadas com critério, tendo seus riscos mitigados³⁵.

A LGPD prevê o tratamento de dados pessoais (art. 7º) e de dados sensíveis (art. 11) em estudos feitos por órgãos de pesquisa, desde que seja garantida a

³³ O Código de Ética Médica define em: Princípios Fundamentais, inciso XI – “O médico guardará sigilo a respeito das informações de que detenha conhecimento no desempenho de suas funções, com exceção dos casos previstos em lei e X – Documentos Médicos; Art. 88. Negar, ao paciente, acesso a seu prontuário, deixar de lhe fornecer cópia quando solicitada, bem como deixar de lhe dar explicações necessárias à sua compreensão, salvo quando ocasionarem riscos ao próprio paciente ou a terceiros”.

³⁴ ALMEIDA, Betânia de Araújo. Ética em pesquisa e dados sensíveis: o que muda? In: SILVA, Angélica Baptista; CUNHA, Francisco José Aragão Pedrosa (orgs.). **Lei geral de proteção de dados e o controle social da saúde**. Porto Alegre: Rede Unida, 2023, p. 280-291.

³⁵ BARRETO, Maurício Lima; ALMEIDA, Bethânia de Araújo; DONEDA, Danilo. Uso e Proteção de Dados Pessoais na Pesquisa Científica. **RDU**, Porto Alegre, v. 16, n. 90, p. 179-194, 2019.

anonimização dos dados pelos agentes de tratamento. Prevê-se, também, o desenvolvimento de estudos em saúde pública (art. 13) que devem seguir alguns cuidados específicos: os dados pessoais devem ser tratados exclusivamente no órgão de pesquisa, restritos à finalidade da pesquisa; além disso, devem ser mantidos em ambiente controlado e seguro, sendo aplicada a anonimização ou pseudoanonimização sempre que possível, de acordo com as normas éticas em pesquisa.

A pesquisa, por realizar o tratamento de dados pessoais, deve buscar o equilíbrio entre proteger o direito dos titulares de dados e garantir o interesse público a partir de métodos que protejam esses dados com o mínimo de processamento para alcançar os objetivos da pesquisa, reduzindo o risco de sua inutilização. Uma vez que a ética em pesquisa faz parte da legalidade do tratamento de dados, o Sistema CEP/CONEP e a Autoridade Nacional de Proteção de Dados (ANPD) devem estabelecer, em conjunto com a comunidade científica, as normas necessárias para a salvaguarda da dignidade, os direitos e a segurança dos participantes em pesquisas cuja fonte de informação são bancos de dados administrativos³⁶.

Em junho de 2023, a ANPD publicou o guia orientativo intitulado “Tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas”, como resultado de um estudo técnico que apresenta boas práticas no tratamento de dados pessoais em pesquisa como orientação aos agentes de tratamento³⁷.

Conforme o guia, o regime jurídico do tratamento de dados pessoais para fins acadêmicos e de pesquisas tem maior flexibilidade, sem prejuízo à liberdade de expressão, à liberdade acadêmica e à promoção da inovação científica e tecnológica, desde que sejam tomadas medidas adequadas de proteção de dados dos titulares e seja seguido o princípio da boa-fé³⁸, conforme previsto no art. 6º da LGPD³⁹.

5. ANONIMIZAÇÃO E PSEUDOANONIMIZAÇÃO DE BASES DE DADOS

Dado pessoal e dado anônimo são opostos na medida em que o primeiro permite a identificação de um indivíduo e o segundo busca ocultar suas características, impedindo sua identificação. A partir de técnicas de anonimização (supressão, generalização, randomização e pseudoanonimização), um dado pessoal pode

³⁶ BARRETO; ALMEIDA; DONEDA, 2019, *op. cit.*

³⁷ VARGAS, Andressa Giroto, *et al.* **Guia Orientativo: tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas.** Brasília, DF: Autoridade Nacional de Proteção de Dados, 2023.

³⁸ I. finalidade, II. adequação, III. necessidade, IV. livre acesso, V. qualidade dos dados, VI. transparência, VII. segurança, VIII. prevenção, IX. não discriminação e X. responsabilização e prestação de contas.

³⁹ VARGAS *et al.*, *op. cit.*

tornar-se um dado anônimo, sendo essa uma das formas de assegurar a confidencialidade desse dado⁴⁰.

O desenvolvimento de técnicas e algoritmos de anonimização em bases secundárias pode reduzir o risco de reidentificação e consequentes ataques cibernéticos. Em tese, um dado anonimizado seria suficiente para garantir a confidencialidade e a proteção dos dados, dispensando o consentimento do usuário e mitigando-se o viés de seleção. Porém, quanto maior o grau de anonimização, mais as informações deixam de fornecer conteúdos relevantes, tornando-se inúteis. Por outro lado, se forem mantidos os dados em sua integridade, o risco de reidentificação cresce substancialmente⁴¹.

Em uma revisão sistemática que avaliou a capacidade de anonimização de 75 (setenta e cinco) algoritmos demográficos, 18 (dezoito) algoritmos de diagnóstico médico e 3 (três) sobre dados genéticos, os autores concluíram que combinação de diferentes técnicas e a aplicação adequada dos algoritmos podem reduzir o risco de reidentificação. No entanto, algoritmos mais eficientes para diagnóstico médico e dados genéticos precisam ser desenvolvidos⁴².

O processo de anonimização envolve uma série de riscos que o agente de tratamento deve considerar⁴³:

- a) grandes volumes de dados têm maiores chances de estabelecerem o processo reverso da identificação, dado o aumento das chances de acerto das associações. Por isso, devem ser aplicadas técnicas mais robustas;
- b) a natureza dos dados define o quanto um dado pode proporcionar informações valiosas, como, por exemplo, informações financeiras e de consumo podem ser mais atrativas por identificarem possíveis vítimas de extorsão;
- c) a cadeia da atividade de tratamento de dados, na qual quanto maior o número de sujeitos que compartilham ou geram a base de dados, maior o risco de reidentificação;
- d) o gerenciamento de identidades e a segmentação referem-se ao controle do acesso dos atores aos dados, reduzindo-se as chances de que pessoas capazes unam as informações, reconheçam um indivíduo e monitorem as atividades dos agentes de tratamento;

⁴⁰ BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. In: DONEDA, Danilo.; MENDES, Laura Schertel Mendes; CUEVA, Ricardo Villas Boas, (orgs.). **Lei geral de proteção de dados (Lei nº 13.709/2018): a caminho da efetividade: contribuições para a implementação da LGPD**. São Paulo: Revista dos Tribunais, 2020, p. 39-54.

⁴¹ SEPAS, Ali, *et al.* Algorithms to anonymize structured medical and healthcare data: A systematic review. *Frontiers in Bioinformatics*, v. 2, 2022, p. 1-11.

⁴² *Ibid.*

⁴³ BIONI, *op. cit.*, p. 39-54.

- e) a inclusão de cláusulas contratuais que delimitem as atividades dos agentes de tratamento ou proíbam o processo de anonimização; e
- f) as técnicas de anonimização devem ser atualizadas continuamente para que não se tornem defasadas e para que as bases não fiquem vulneráveis a técnicas novas de reidentificação.

No §4º do art. 13 da LGPD, a pseudoanonimização é definida como um tratamento do dado pessoal de tal forma que não há possibilidade de associação, direta ou indiretamente, a um indivíduo, mas que o agente de tratamento pode, por meio de informação adicional e em ambiente seguro e controlado, reverter suas características originais que caracterizam determinado indivíduo. O uso da pseudoanonimização está restrito a estudos de saúde pública.

A aplicação de técnicas de anonimização requer conhecimento técnico sofisticado, especializado e, a depender do grau de complexidade da base de dados, implica custos mais elevados. Em alguns países foram criados centros de dados orientados para pesquisa que desenvolvem espaços com infraestrutura adequada e profissionais especializados, além de elaborarem políticas de governança de dados com objetivo de prover acesso a dados de qualidade, de forma segura e controlada para pesquisas, avaliações e informações que subsidiem políticas públicas.

Geralmente, são organizações em parceria com o governo, universidades e instituições de pesquisa que agregam grandes bases de dados de registros administrativos que, apesar de não terem sido coletadas para fins de pesquisa, podem ser utilizadas em projetos de pesquisa. Tais centros, após o tratamento de dados e a desidentificação dos indivíduos, viabilizam o acesso integrado de bases de dados às pesquisas científicas, buscando um equilíbrio entre o desenvolvimento da ciência para o benefício coletivo e a garantia da proteção dos dados pessoais.

Cite-se, ainda, a iniciativa brasileira promovida pela Fiocruz, que é o Centro de Integração de Dados e Conhecimentos para a Saúde (CIDACS)⁴⁴.

6. CONCLUSÕES

Na área da pesquisa em saúde, a LGPD instigou os atores envolvidos no tratamento de dados a revisitarem os fluxos de informações e, a partir daí, identificarem os pontos de vulnerabilidade e possíveis falhas na garantia da privacidade dos indivíduos e na proteção dos dados pessoais.

Em um primeiro momento, sua aplicação trouxe dúvidas que levaram muitos órgãos de pesquisa e comitês de ética em pesquisa a recusarem a utilização de bases de dados secundárias em razão da impossibilidade de consentimento dos titulares.

⁴⁴ BARRETO; ALMEIDA; DONEDA, 2019, *op. cit.*

No entanto, com o passar do tempo e uma melhor compreensão da LGPD, os órgãos de pesquisa buscaram se adequar e atingir o equilíbrio entre a proteção dos dados pessoais e o desenvolvimento de pesquisas científicas, sem prejuízo do potencial que as bases de dados em saúde podem fornecer para o desenvolvimento científico, tecnológico e de inovação, além das garantias de liberdade de expressão e pluralismo de ideias, bem como da liberdade acadêmica.

A aplicação de técnicas de anonimização por meio do desenvolvimento de algoritmos é uma possibilidade para a garantia da proteção dos dados. Porém, deve-se considerar os custos de desenvolvimento dessas ferramentas e a necessidade de profissionais especializados e em constante atualização, uma vez que, a cada dia, os dados se tornam ativos valiosos, aumentando as chances de reidentificação e, conseqüentemente, de ataques cibernéticos que podem levar a constrangimentos, à exposição de dados pessoais ou a situações mais graves como extorsões e roubos.

A criação de centros de dados orientados para pesquisa pode ser uma opção, cenário no qual se concentrariam as condições necessárias para o desenvolvimento de mecanismos de anonimização, com o controle dos dados pessoais de forma a garantir sua proteção.

Atualmente, muitos dados estão restritos às mãos de um número limitado de empresas e instituições (públicas e privadas) que podem direcionar o uso de dados para seu próprio interesse, o qual nem sempre estará em consonância com as necessidades da sociedade. O compartilhamento de dados abertos, desde que garanta a privacidade dos sujeitos e proteção dos dados pessoais, deve ser incentivado para que possam ser disponibilizados para o maior número de pesquisadores e órgãos de pesquisa, buscando-se ampliar o conhecimento, reduzir a concentração de poder e beneficiar a sociedade com ética e responsabilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAMOVAY, Ricardo; ZANATTA, Rafael Augusto Ferreira. Dados Pessoais Abertos: Pilares dos Novos Mercados Digitais? **RDU**, Porto Alegre, v. 16, n. 90, p. 155-178, 2019.

ALMEIDA, Betânia de Araújo. Ética em pesquisa e dados sensíveis: o que muda? In: SILVA, Angélica Baptista; CUNHA, Francisco José Aragão Pedrosa (orgs.). **Lei geral de proteção de dados e o controle social da saúde**. Porto Alegre: Rede Unida, 2023, p. 280-291.

ALMEIDA, Betânia de Araújo.; PIMENTA, Denise Moraes; BARRETO, Mauricio Lima. Compartilhamento e abertura de dados na Ciência Aberta e a pesquisa em saúde pública na Lei Geral de Proteção de Dados. **Bahia Análise & Dados**, Salvador, v. 31, n. 2, p. 78-92, 2021. ALMEIDA FILHO, Naomar de. Bases históricas da Epidemiologia. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 2, n. 3, p. 304-311, set. 1986.

BARRETO, Maurício Lima; ALMEIDA, Bethânia de Araújo; DONEDA, Danilo. Uso e Proteção de Dados Pessoais na Pesquisa Científica. **RDU**, Porto Alegre, v. 16, n. 90, p. 179-194, 2019.

BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. In: DONEDA, Danilo.; MENDES, Laura Schertel Mendes; CUEVA, Ricardo Villas Bôas, (orgs.). **Lei geral de proteção de dados (Lei nº 13.709/2018): a caminho da efetividade: contribuições para a implementação da LGPD**. São Paulo: Revista dos Tribunais, 2020, p. 39-54.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 3 mar. 2024.

BRASIL. Emenda Constitucional nº 115, de 10 de fevereiro de 2022. Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais. **Diário Oficial da União**, Brasília, 2022.

BRASIL. Lei nº 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências. **Diário Oficial da União**, Brasília, DF, 1990. BRASIL. Lei nº 10.406, de 10 de janeiro de 2002. Institui o Código Civil. **Diário Oficial da União**, Brasília, DF, 10 jan. 2002.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 17 jun. 2024.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução nº 466, de 12 de dezembro de 2012**. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília, DF: Ministério da Saúde, 2012.

BRASIL. Conselho Nacional de Saúde. **Resolução nº 588, de 12 de julho de 2018**. Institui a Política Nacional de Vigilância em Saúde (PNVS). Brasília, DF: Conselho Nacional de Saúde, 2018. Disponível em: <https://conselho.saude.gov.br/resolucoes/2018/Reso588.pdf>. Acesso em 3 abr. 2026.

BRASIL. Ministério da Saúde. **Sistemas de informação em saúde**. 2024. Disponível em: <https://www.gov.br/aids/pt-br/svsa-na-gestao-municipal/agrivos-e-doencas-cronicas-nao-transmissiveis/links-de-interesse-sobre-agrivos-e-doencas-cronicas-nao-transmissiveis/sistemas-de-informacao-em-saude>. Acesso em: 3 mar. 2026.

BRASIL. Ministério da Saúde. **Portaria nº 188, de 3 de fevereiro de 2020**. Declara Emergência em Saúde Pública de importância Nacional (ESPIN) em decorrência da Infecção Humana pelo novo Coronavírus (2019-nCoV). Brasília, DF: Ministério da Saúde, 2020.

CONSELHO FEDERAL DE MEDICINA. Resolução nº 1.638/2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. **Diário Oficial da União**, Brasília, 2002.

DALLARI, Analluza Bolivar. Enquadramento do contrato de pesquisa clínica à luz das normas de aplicação imediata brasileiras. In: DALLARI, Analluza Bolivar (org.). **Contrato de Pesquisa Clínica: aspectos práticos e jurídicos**. São Paulo: Revista dos Tribunais, 2019, p. 163-2010.

DONEDA, Danilo. A LGPD como elemento estruturante do modelo brasileiro de proteção de dados. In: **Lei geral de proteção de dados (Lei nº 13.709/2018): a caminho da efetividade: contribuições para a implementação da LGPD**. São Paulo: Thomson Reuters; Revista dos Tribunais, 2020, p. 243-255.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. 3. ed. São Paulo: Thomson Reuters Brasil, 2021, p. 30-47.

LIMA, Claudia Risso de Araujo *et al.* Departamento de Informática do SUS – DATASUS: a experiência de disseminação de informações em saúde. In: BRASIL (org.). **A experiência brasileira em sistemas de informação em saúde**. Brasília: Editora do Ministério da Saúde, 2009. v. 1. p. 109-128.

LUBARSKY, Boris. Reidentificação de dados “anonimizados”. **Georgetown Law Technology Review**, 2017. Disponível em: https://georgetownlawtechreview-org.translate.google/re-identification-of-anonymized-data/GLTR-04-2017/?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc. Acesso em: 3 abr. 2026.

PAIVA, Natália Santana, *et al.* Sistema de informações sobre nascidos vivos: um estudo de revisão. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 16, supl. 1, p. 1211-1220, 2011.

SALSBURG, David. **Uma senhora toma chá...**: Como a estatística revolucionou a ciência no século XX. São Paulo: Zahar, 2009.

SEPAS, Ali, *et al.* Algorithms to anonymize structured medical and healthcare data: A systematic review. **Frontiers in Bioinformatics**, Lausanne, v. 2, 2022, p. 1-11.

SNOW, John. **Sobre a maneira de transmissão do cólera**. São Paulo: Hucitec, 1999.

VARGAS, Andressa Giroto, *et al.* **Guia Orientativo: tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas**. Brasília, DF: Autoridade Nacional de Proteção de Dados, 2023.

WARREN, Samuel; BRANDEIS, Louis. The right to privacy. **Harvard Law Review**, Cambridge, v. IV, n. 15, dez. 1890. Disponível em: https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html. Acesso em: 3 abr. 2026.

Submetido em: 23/01/2025

Aprovado em: 22/01/2026