

ANÁLISE DE DADOS APLICADA À AVALIAÇÃO PRELIMINAR DE RISCO DA OCORRÊNCIA DE FÁRMACOS NOS ESGOTOS DAS CAPITAIS DO BRASIL

Simão Pedro de Oliveira Urpia¹

Karla Patrícia Santos Oliveira Rodriguez Esquerre²

Luciano Matos Queiroz³

Júlia Carolina Braz de Freitas Bijos⁴

Resumo

Com o aumento do consumo de medicamentos, especialmente em centros urbanos, surgiram preocupações quanto ao descarte inadequado e a excreção desses compostos, que ao atingir o meio ambiente, prejudicam os ecossistemas onde foram despejados e, potencialmente, a saúde humana, principalmente, por contribuir para o desenvolvimento de resistência antimicrobiana. O objetivo maior desta pesquisa foi investigar o risco ambiental dos resíduos de fármacos nos esgotos das capitais brasileiras, analisando suas concentrações e identificando os medicamentos mais críticos para o controle ambiental. A pesquisa buscou apoiar estratégias de monitoramento e mitigação de riscos associados ao descarte de medicamentos. Para tanto, a metodologia incluiu a análise de dados de vendas de medicamentos controlados, obtidos do Sistema Nacional de Gerenciamento de Produtos Controlados (SNGPC) da Anvisa, bem como o tratamento desses dados e processamento no R. A partir disso, foi possível identificar os fármacos e os locais do país que correm maior risco ambiental, de modo a contribuir para a definição de diretrizes que possam auxiliar em uma regulamentação específica para esses poluentes no Brasil. O trabalho foi capaz de demonstrar a importância de um monitoramento e de estratégias de tratamento de esgoto mais eficazes na remoção de resíduos farmacêuticos, visando à mitigação dos riscos à saúde ambiental e promovendo uma abordagem preventiva para a proteção dos ecossistemas aquáticos, além de evidenciar as vantagens de utilizar a ciência de dados para a solução de problemas ambientais complexos.

Palavras-chave: risco ambiental; fármacos - dados; R; concentração ambiental prevista (CAP).

Abstract

With the increase in medication consumption, especially in urban centers, concerns have arisen regarding the improper disposal and excretion of these compounds. When they reach the environment, they harm ecosystems where they are deposited and potentially human health, primarily by contributing to the development

¹Universidade Federal da Bahia (UFBA), urpiasimao@gmail.com

²Universidade Federal da Bahia (UFBA), karla.esquerre@gmail.com

³Universidade Federal da Bahia (UFBA), lmqueiroz@ufba.br

⁴Universidade Federal da Bahia (UFBA), julia.bijos25@gmail.com

of antimicrobial resistance. The main objective of this research was to investigate the environmental risk of pharmaceutical residues in wastewater systems of Brazilian capitals, analyzing their concentrations and identifying the most critical drugs for environmental control. The study aimed to support monitoring strategies and risk mitigation related to drug disposal. To this end, the methodology included the analysis of controlled drug sales data obtained from the National System for the Management of Controlled Products (SNGPC) of Anvisa, as well as the processing of these data using R. This allowed for the identification of pharmaceuticals and locations in the country at higher environmental risk, contributing to the development of guidelines that could aid in the regulation of these pollutants in Brazil. The study demonstrated the importance of monitoring and more effective wastewater treatment strategies for the removal of pharmaceutical residues, aiming at mitigating risks to environmental health and promoting a preventive approach to the protection of aquatic ecosystems. It also highlighted the advantages of using data science to solve complex environmental issues. **Keywords:** environmental risk, drugs – data, R, predicted environmental concentration (PAC).

Introdução

De acordo com a Organização das Nações Unidas para a Educação, Ciência e Cultura (Unesco, 2023),

“poluentes emergentes em um sentido amplo são qualquer substância química sintética ou natural ou qualquer microrganismo que não é comumente monitorado ou regulado no meio ambiente, com efeitos adversos potenciais conhecidos ou suspeitos na ecologia ou na saúde humana”.

Constata-se, portanto, que a ausência de monitoramento adequado desses contaminantes é um importante gargalo para definição de estratégias que visem mitigar seus impactos adversos.

A contaminação direta por esses compostos é causada, por exemplo, por efluentes da indústria farmacêutica gerados nas operações de limpeza ou no descarte inadequado de produtos fora de padrões de qualidade. Já a contaminação indireta, considerada a principal via de contaminação ambiental, pode ocorrer nos esgotos sanitários e provém, principalmente, da excreção e descarte inadequado de medicamentos – com prazo de validade vencido (Sales et al., 2012). Portanto, está diretamente relacionada à cadeia de produção, ao consumo e ao ciclo de vida dessas substâncias.

Desse modo, surge um questionamento pertinente: como investigar o potencial impacto negativo desses poluentes nos diferentes compartimentos ambientais, principalmente nos grandes centros urbanos? Monitorar a qualidade da água de corpos hídricos, sedimentos e do solo; efluentes de Estações de Tratamento de Esgoto (ETEs); água de abastecimento; e

padrões de consumo, produção e descarte desses compostos costuma ser uma tarefa altamente complexa. São imprescindíveis campanhas de monitoramento baseadas em coletas de amostras ao longo do tempo para que os dados sejam confiáveis, e essas investigações demandam tempo e consomem muitos recursos financeiros.

A aplicação de ferramentas da Ciência de Dados, aliada a metodologias de análise de risco, pode aprofundar a compreensão dos fatores que contribuem para a ampla ocorrência de fármacos no meio ambiente, pode complementar métodos de monitoramento baseados em campanhas amostrais. Estas ferramentas, habilitam a identificação de cidades com necessidade de monitoramento prioritário, bem como dos medicamentos que podem estar presentes em maior quantidade no ambiente. Neste contexto, o banco de dados Sistema Nacional de Gerenciamento de Produtos Controlados (SNGPC) foi selecionado como fonte de dados de comercialização de fármacos em municípios brasileiros. Esse banco marca a transição da escrituração tradicional em livro físico para a escrituração eletrônica, exigindo a transmissão dos dados à Anvisa. Contudo, consoante à Resolução da Diretoria Colegiada (RDC) nº 586/2021, apesar da necessidade de manutenção dos registros internos pelos estabelecimentos, a partir de 5 de outubro de 2021 a obrigatoriedade da transmissão de dados à Anvisa foi revogada. O registro foi suspenso a partir de 23 de dezembro de 2022.

Objetivo

Este trabalho buscou aplicar ferramentas para análise de dados da comercialização de dez fármacos e avaliar o risco da possível ocorrência nos esgotos das capitais do Brasil

Material e Método

Inicialmente realizou-se um extenso tratamento dos dados extraídos do SNGPC. Os dados do SNGPC estão disponibilizados online no site do governo federal e são armazenados em 82 arquivos no formato.csv. Cada arquivo condensa os dados escriturados referentes a um mês dos sete anos em que o registro ocorreu de forma obrigatória 2014 a 2021. Esses bancos de dados estão originalmente divididos em 15 campos, com os seguintes títulos: “ANO_VENDA”, “MES_VENDA”, “UF_VENDA”, “MUNICIPIO_VENDA”, “PRINCIPIO_ATIVO”, “DESCRICAO_APRESENTACAO”, “QTD_VENDIDA”, “UNIDADE_MEDIDA”, “CONSELHO_PRESCRITOR”, “TIPO_RECEITUARIO”, “UF_CONSELHO_PRESCRITOR”, “CID10”, “SEXO”, “IDADE” e “UNIDADE_IDADE” (Agência Nacional de Vigilância Sanitária, 2020).

A leitura dos arquivos foi realizada com as funções `read.csv()` e `read_excel()`, funções que transformam os arquivos tabulares em objetos `data.frame`, estrutura fundamental da linguagem R para manipulação de dados.

Durante a execução do estudo, procedimentos de manipulação de dados, geração de representações gráficas, análise exploratória e computação de cálculos estatísticos foram

realizados utilizando-se um conjunto de ferramentas computacionais especializadas. Dentre elas, destacam-se o software Microsoft Excel® e a linguagem de programação estatística R Core Team® (2024), através de seu Ambiente de Desenvolvimento Integrado (IDE) RStudio Team® (2024). A escolha do R foi feita por esta ser uma plataforma extremamente eficaz para o tratamento e análise de grandes volumes de dados complexos.

Para otimizar a manipulação e visualização dos dados, foi realizada uma série de operações nos bancos de dados, focando na redução do tamanho sem comprometer informações relevantes. Foi feita a conversão para o formato binário “.RData” com as funções `save()`, `saveRDS()` e `load()`, as quais permitem armazenar e recuperar objetos com maior velocidade e ocupam cerca de 23 vezes menos espaço na memória do dispositivo que os mesmos bancos em .csv. Para a construção do banco optou-se pelos dez princípios ativos mais vendidos durante o período de 2014 a 2019, e as 27 capitais dos estados brasileiros foram selecionadas para a investigação.

O pacote `dataMaid` foi utilizado para gerar relatórios automatizados de qualidade dos dados, identificando variáveis com alta proporção de dados ausentes, valores discrepantes e problemas de codificação. Neste contexto, identificou-se a necessidade de realizar o tratamento dos dados aberrantes presentes na coluna “QTD_VENDIDA”, que registra a quantidade vendida de medicamentos em caixas ou frascos. A coluna apresentou um número elevado de outliers, identificados por meio do Método do Intervalo Interquartil (IQR), os quais foram inicialmente marcados como “NA”. Como essa variável é fundamental para o cálculo da massa total de medicamentos vendidos, optou-se por realizar a imputação desses valores. Embora a distribuição da variável apresente certo grau de assimetria, observou-se que a concentração de dados é muito elevada em torno de um único valor, o que fez com que a média e a mediana fossem praticamente idênticas, situando-se entre 2 e 2,5. Considerando que os dados dessa coluna são números naturais, a substituição dos valores ausentes por “2” foi adotada, pois representa o valor inteiro mais próximo das medidas de tendência central observadas. Essa abordagem permitiu preservar a representatividade da amostra e a precisão das análises subsequentes.

Dentre os pacotes utilizados, destaca-se os disponíveis no `tidyverse`, um conjunto coerente de pacotes para ciência de dados (como `dplyr`, `ggplot2`, `readr`, `tibble`), e o `janitor`, voltado à limpeza de dados e padronização de colunas. O pacote `data.table` foi empregado para manipulação eficiente dos conjuntos de dados, se destacando por reduzir o uso de memória e tempo de processamento.

A extração de informações textuais relevantes, como concentração (dosagem do fármaco), volume e tipo de embalagem, foi viabilizada por expressões regulares aplicadas por funções como `str_extract()` e `gsub()`, que utilizam a biblioteca `stringr`. Essa capacidade torna o R particularmente adequado para tarefas de mineração de texto e normalização de campos despadronizados, como a descrição das formas de apresentação comercializadas.

Devido à complexidade das informações em formato de texto, a mineração de texto

foi revisada, utilizando o Excel, e passou por ajustes feitos diretamente nas strings que estavam fora dos padrões. A exportação para arquivos Excel foi feita com o pacote writexl, por meio da função write_xlsx(), facilitando a integração dos recursos estatísticos do Excel com os do R e também facilitando o uso de tabelas em formatação compatível com a divulgação científica.

As análises estatísticas descritivas foram realizadas com as funções group_by(), summarise(), n(), mean() e sum() do pacote dplyr e a visualização dos dados foi conduzida com o pacote ggplot2. Com isso, foram criados gráficos de barras, mapas de calor e representações temporais com geom_col(), geom_tile() e gganimate, personalizadas por theme(), labs() e escalas específicas.

Por fim, com base no modelo proposto por Silva (2019), realizou-se o cálculo da Concentração Ambiental Prevista (CAP) usando a Equação 1 para cada uma das cidades, levando em consideração as informações do banco de dados. Para a obtenção dos resultados, contudo, realizaram-se algumas adequações no método proposto pelo autor, que se refletiram no equacionamento: (i) a fração de esgoto tratada no município em relação à água consumida representa a capacidade local de impedir que esses fármacos alcancem as águas de superfície no estado in natura; (ii) todo o fármaco consumido é excretado de forma inalterada na rede municipal de esgotos e esse serviço atende 100% dos domicílios; (iii) os dados foram modificados para valores anuais conforme a Equação 1; (iv) utilizaram-se dados de venda de medicamentos ao invés dos dados de consumo, ou seja, considerou-se que não há desperdício de medicamentos; (v) No Brasil, cada pessoa gera, em média, 150 litros de esgoto por dia.

$$CAP = VFA/QE + QFE * 10 \text{ Eq. (1)}$$

Onde:

CAP: Concentração Ambiental Prevista, em $\mu g \cdot L^{-1}$;

VFA: Volume de fármaco anual vendido, $\mu g \cdot ano^{-1}$;

QE: Vazão do esgoto doméstico, em $L \cdot ano^{-1}$;

QFE: Vazão do effluente da fonte emissora, em $L \cdot ano^{-1}$;

10: Fator de diluição adotado.

A QE é calculada a partir do consumo de água per capita da população atendida pela rede de esgotos, utilizando-se a equação 2:

$$QE = \text{Volume água consumida per capita/dia} * \text{Coeficiente de retorno} * \text{População atendida} * 365 \text{ Eq. (2)}$$

A depender da norma considerada, existem diferentes limites de CAP considerados seguros. Um destes é o proposto pela norma da União Europeia (EMEA,2006), de $0,01 \mu g \cdot L^{-1}$, usado comumente para dados de esgotos europeus. No entanto, este estudo tomou como base o limite adotado por Silva (2019), por ser aplicado a dados brasileiros,

no qual se o valor da CAP for inferior a $0,1\mu g \cdot L^{-1}$, é improvável que o fármaco represente um risco para o meio ambiente. Se o valor for igual ou superior a $0,1\mu g \cdot L^{-1}$, a avaliação deve prosseguir para as etapas seguintes.

Resultados e Discussão

Neste estudo, um grande desafio enfrentado foi o de manipular e visualizar o grande volume de dados. Cada arquivo .csv continha inicialmente cerca de 6.000.000 (seis milhões) de linhas e, portanto, 90.000.000 (noventa milhões) de dados.

Considerando que o metabolismo e a excreção originada de seres humanos são os principais contribuintes para a presença de fármacos em efluentes, definiu-se um recorte específico que incluiu dados de vendas de todas as 27 capitais do país. A escolha de focar nas capitais, em vez de selecionar as cidades mais populosas do país, oferece uma vantagem significativa: a diversidade dos dados. Mesmo após a aplicação do filtro, a amostragem ainda abrange os diversos perfis de consumo e o ciclo de vida dos medicamentos comercializados em todo o território nacional.

Devido à complexidade das informações presentes no banco de dados, foi necessário fazer algumas considerações. Nesse sentido, os medicamentos com mais de um princípio ativo foram tratados como possuindo apenas o princípio ativo principal. Assim, a coluna “PRINCIPIO_ATIVO” foi usada como base para extrair a informação referente ao primeiro princípio ativo apresentado. Esse registro foi, então, armazenado na coluna “PRINCIPIO_ATIVO_PRINCIPAL”.

Além disso, para a construção do banco optou-se pelos dez princípios ativos mais vendidos durante o período analisado. É importante salientar que o ano de 2020 também foi marcado pelo início da pandemia causada pelo covid-19, que provocou um comportamento extremamente atípico na indústria e no mercado farmacêutico. Essa alteração regulatória e a ocorrência da pandemia resultaram em uma base de dados mais confiável para o período entre 2014 e 2019, de tal modo que o banco foi filtrado para apenas esse período histórico.

No banco de dados unificado, uma nova coluna denominada “MASSA_TOTAL” foi criada, com o objetivo de armazenar a massa total de princípio ativo vendido para cada medicamento em cada ocorrência de venda. Para cada medicamento, a massa total foi calculada através da multiplicação do volume por unidade, do número de unidades por embalagem, da dosagem por unidade e da quantidade de embalagens vendidas. O cálculo de variáveis derivadas, como massa total (mg) por produto, foi realizado com operadores vetoriais (*), que aplicam multiplicações linha a linha sobre os vetores envolvidos. A partir da coluna “MASSA_TOTAL” foi possível identificar os dez princípios ativos de antibióticos mais vendidos nas capitais brasileiras (Tabela 1). Assim, esses princípios ativos são, potencialmente, os que apresentam maior risco ambiental e à saúde humana.

Portanto, para a realidade brasileira, deveriam ser considerados componentes prioritários em listas de monitoramento e controle ambiental.

Tabela 1: Massa de vendas totais para cada princípio ativo nas 27 capitais brasileiras entre os anos de 2014 e 2019

Princípio ativo	Massa acumulada descartada (ton)
Amoxicilina	26,5
Cefalexina	4,2
Cloridrato de Ciprofloxacino	3,5
Azitromicina	2,6
Metronidazol	2,2
Levofloxacino	2,2
Dexametasona	1,8
Sulfametoxazol	1,8
Norfloxacino	0,8
Cloridrato de Moxifloxacino	0,2
Fonte: elaborada pelo autor (2024)	

O principal destaque observado é a amoxicilina, que lidera a lista com um total de 26,5 toneladas vendidas, demonstrando um volume de vendas significativamente maior em relação aos demais princípios ativos.

Em estudo que buscou realizar uma avaliação da qualidade das prescrições de antimicrobianos dispensadas em unidades públicas de saúde de Belo Horizonte em 2002, observou-se que no período estabelecido foram prescritos 19 antibacterianos distintos. O antibiótico prescrito com maior frequência foi a amoxicilina – 49,7% nas receitas internas e 66,3% nas externas (Abrantes *et al.*, 2007). Assim como o comportamento observado nos dados da Tabela 1, a amoxicilina também é indicada como antibiótico mais vendido/consumido em outras pesquisas como, Berquó *et al.* (2004), Castro *et al.* (2020) e Marliere *et al.* (2000).

Esse é o princípio ativo mais vendido do banco de dados, em todas as capitais do país, em todos os anos. Isso reflete seu uso disseminado como antibiótico de amplo espectro, uso primário, baixo custo, com boa adaptação às bactérias desde sua formulação, e comumente receitada em conjunto com outros medicamentos, fato frequentemente discutido na área médica (White *et al.* (2004); Huttner *et al.* (2020)).

Com o objetivo de analisar de forma mais detalhada as tendências e variações nas vendas de medicamentos por cidade e por substância, foram elaborados os gráficos apresentados nas Figuras 1, 2 e 3. Cada gráfico exibe visualmente os dados de venda total, em quilogramas, no período de 2014 a 2019. As Figuras 1 e 2 também incluem as equações das retas ajustadas por meio de modelos de regressão linear simples, aplicados às

variáveis MASSA_TOTAL_kg e ANO_VENDA. Para medicamentos ou cidades não incluídos no corpo principal do texto, os resultados encontram-se no anexo. A interpretação das tendências representadas nos gráficos pode ser realizada com base na inclinação da reta (coeficiente angular), no coeficiente de determinação (R^2), que expressa a proporção da variabilidade explicada pelo modelo, e no p-valor do teste t de Student, que indica a significância estatística do coeficiente de regressão.

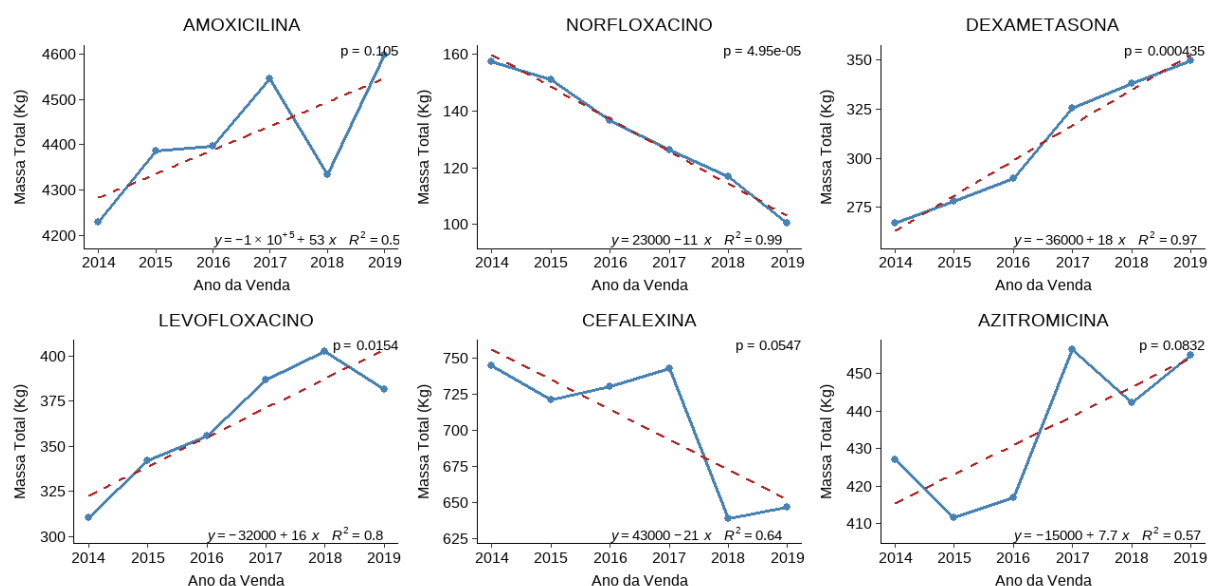


Figura 1: Tendência de vendas de medicamentos controlados por ano: Amoxicilina, Azitromicina, Norfloxacin, Levofloxacin, Cefalexina, Dexametasona

Fonte: elaborada pelo autor (2024)

Dentre os dez princípios ativos analisados, apenas a Dexametasona, o Levofloxacin e o Norfloxacin apresentaram modelos de regressão linear simples com p-valores do teste t de Student inferiores a 0,05, indicando uma associação estatisticamente significativa entre o tempo e o volume total vendido. Além disso, os modelos correspondentes apresentaram bom ajuste, conforme evidenciado pelos coeficientes de determinação (R^2) elevados. Assim, para esses princípios ativos, é viável realizar projeções de consumo para os anos subsequentes com base nas tendências observadas. No caso da Dexametasona e do Levofloxacin, verificou-se uma tendência de crescimento no consumo anual estimado em aproximadamente 30% e 20%, respectivamente. Em contrapartida, o Norfloxacin apresentou tendência de redução, com uma queda anual estimada de cerca de 36%. Em relação à amoxicilina, princípio ativo mais vendido do banco de dados, é possível observar que apesar de não possuir um bom ajuste a reta de tendência linear, indica-se um aumento nas vendas de cerca de 9% ao ano.

Os perfis da comercialização para os medicamentos costumam ter forte correlação com as tendências de uso de sua classe terapêutica. Os autores Van Boeckel *et al.* (2014) indicaram tendências de consumo para diversas classes de antibióticos em nível global,

comparando o consumo nos anos 2000 e 2010, e avaliaram que para as cefalosporinas e fluoroquinolonas, ocorreu um aumento do consumo, observado principalmente em países de renda média, como Índia, China e Brasil, comportamento similar ao observado nas Figuras 1 e 2. Para os autores, esses comportamentos podem ser justificados pelas mudanças na epidemiologia global da resistência aos antibióticos e ao aumento da ocorrência de algumas infecções endêmicas em países em desenvolvimento.

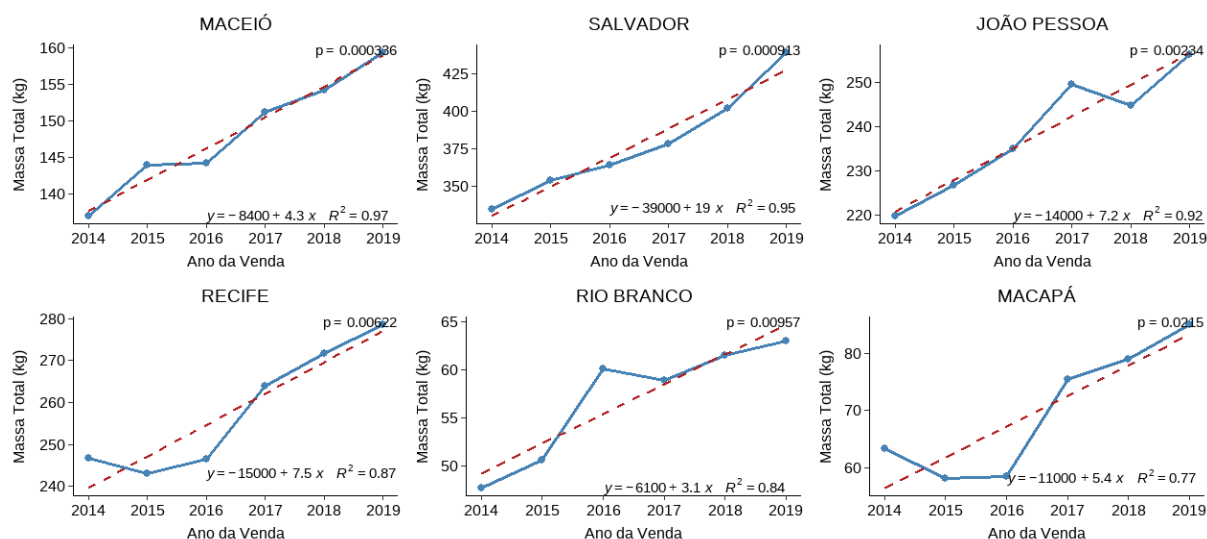


Figura 2: Tendência de vendas de medicamentos controlados por cidade: Maceió, Salvador, João Pessoa, Recife, Ríio Branco, Macapá

Fonte: elaborada pelo autor (2024)

Das 27 capitais analisadas, apenas 10 apresentaram modelos de regressão linear simples com p-valor do teste t de Student inferior a 0,05 para o coeficiente da variável ANO_VENDA, indicando uma associação estatisticamente significativa entre o ano e o consumo de medicamentos. Nesses casos, a qualidade do ajuste também foi satisfatória, com coeficientes de determinação (R^2) superiores a 0,7, sugerindo que a tendência temporal explica uma parcela substancial da variação observada. Assim, é possível considerar a utilização desses modelos para projeções futuras do consumo. Dentre essas capitais, destacam-se as seis cidades apresentadas na Figura 2, cujos modelos apresentaram os melhores ajustes. Nessas localidades, a taxa média de crescimento anual das vendas variou de aproximadamente 3%, como observado em Recife, João Pessoa e Maceió, até cerca de 8,5% em Macapá.

Para casos em que os modelos não representaram um ajuste satisfatório, podem ser aplicadas outras técnicas para melhor compreender o perfil das vendas, como modelos não-lineares ou demais modelos preditivos baseados em aprendizado a partir dos dados.

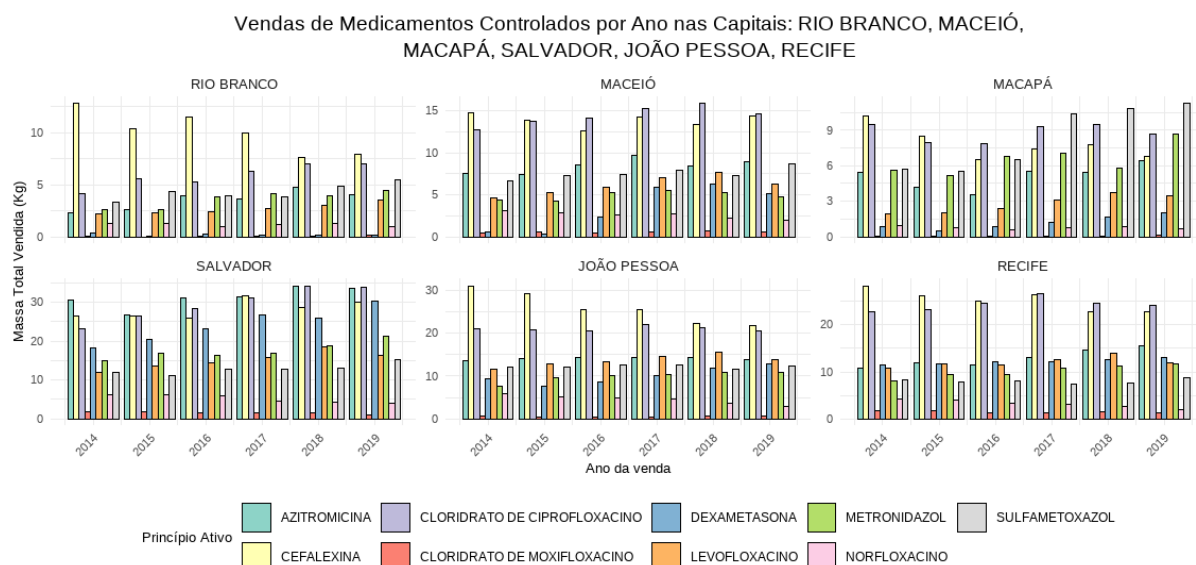


Figura 3: Vendas de medicamentos controlados por cidade e por medicamento: Maceió, Salvador João Pessoa, Recife, Ró Branco, Macapá

Fonte: elaborada pelo autor (2024)

Os gráficos de barras da figura 3, apresentados sem a amoxicilina, facilitam a visualização de picos de consumo e de possíveis comportamentos atípicos dos 9 demais medicamentos, facilitando também a percepção das tendências de cada medicamento em cada cidade. É importante evidenciar que o comportamento geral de um medicamento não se reflete da mesma forma em todas as cidades.

Por fim, a CAP de todas as cidades e medicamentos foram calculadas, mas apenas os medicamentos cujo cálculo do CAP indicou maior risco seguem representados visualmente e discutidos neste artigo.

A tabela 2 mostra a distribuição da CAP de princípios ativos nas 27 capitais brasileiras, entre 2014 e 2019. Observa-se que a maioria dos CAPs (1.451) estão abaixo do limite de $0,1 \mu\text{g.L}^{-1}$, considerado seguro para o ambiente por Silva (2019). No entanto, 169 CAPs ultrapassam esse limite, indicando potencial risco de contaminação.

Considerando o limite de $0,01 \mu\text{g.L}^{-1}$ proposto pela norma da União Europeia (EMEA,2006) nota-se que 1.235 CAPs o ultrapassaram, sugerindo que o mesmo cenário em um país da União Europeia indicaria a necessidade de prosseguir para as próximas etapas da Análise de Risco Ambiental. Apenas 385 CAPs estão abaixo desse limite.

Tabela 2: Panorama de resultados de CAPs calculadas para as 27 capitais do Brasil para os anos de 2014 a 2019, considerando os dois limites de segurança

Categorias	Limite de 0.1 $\mu\text{g.L}^{-1}$	Limite de 0.01 $\mu\text{g.L}^{-1}$
Quantidade de CAPs acima do limite	169	1235
Quantidade de CAPs abaixo do limite	1451	385
Total de CAPs	1620	1620
Fonte: elaborada pelo autor (2024)		

A Figura 4 ilustra a CAP do fármaco com o maior número de valores acima de 0,1 $\mu\text{g.L}^{-1}$ nas 27 capitais brasileiras ao longo dos anos.

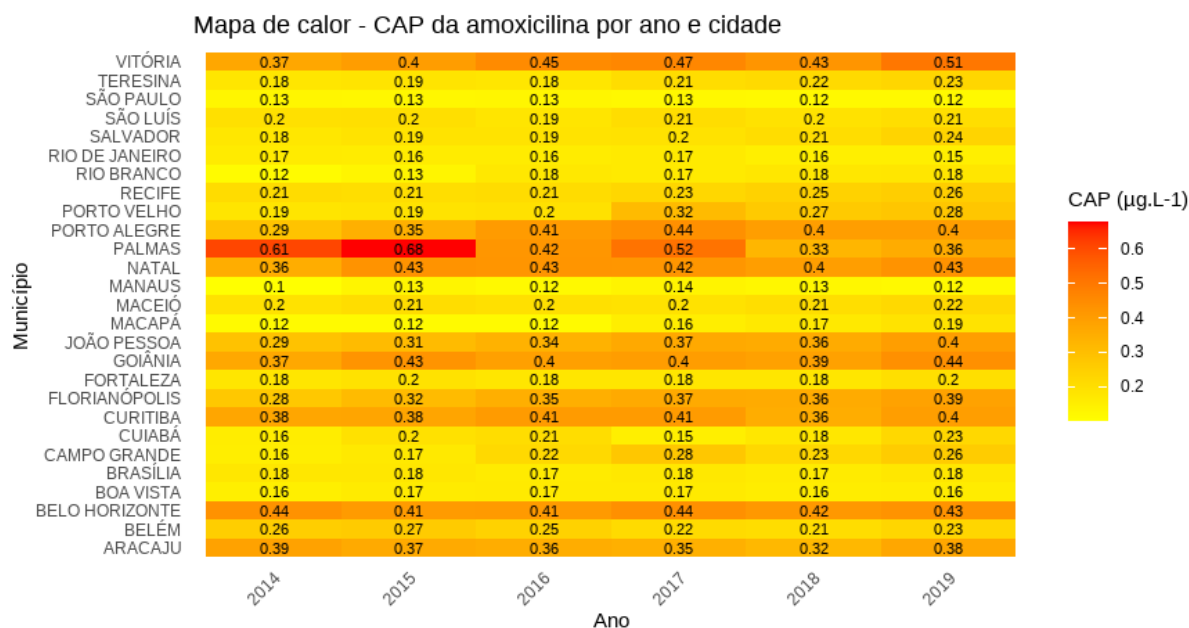


Figura 4: Mapa de calor das CAPs para amoxicilina por cidade e ano

Fonte: elaborada pelo autor (2024)

Em relação à amoxicilina, observa-se que Palmas, capital do Tocantins, apresenta os valores de CAP mais elevados, atingindo pico de $0.68 \mu\text{g} \cdot \text{L}^{-1}$ em 2015. Belo Horizonte e Curitiba também evidenciam valores relativamente altos, atingindo $0.51 \mu\text{g} \cdot \text{L}^{-1}$ e $0.38 \mu\text{g} \cdot \text{L}^{-1}$, respectivamente, em anos diferentes. A CAP da amoxicilina mostra uma tendência crescente na maioria das cidades ao longo dos anos. No entanto, algumas cidades, como Palmas, demonstram uma tendência decrescente após atingirem picos. Em todas as cidades e em todos os anos são observados valores de CAP superiores a $0,1 \mu\text{g} \cdot \text{L}^{-1}$, que podem ter implicações ambientais significativas, incluindo potenciais efeitos

ecotóxicos. Este estudo indica, portanto, que há necessidade de prosseguir com a análise de risco ambiental para a amoxicilina em todas as capitais do país.

É relevante também para a tomada de decisões avaliar os índices de atendimento com tratamento de esgoto para essas capitais. Para a cidade de Natal, que possui anos com o valor de CAP acima de $0.1 \mu g \cdot L^{-1}$ para mais de um medicamento, o atendimento das redes de esgoto ocorre em menos de 50% dos domicílios em 2022, aumentando ainda mais o alerta de controle ambiental. Boa Vista apresenta atendimento das redes de esgoto presente em menos de 62% das residências e possui valores CAP que indicam a necessidade de continuidade para a avaliação de risco. De maneira geral, todas as cidades apresentam a necessidade de prosseguir com as demais etapas para a avaliação de risco ao se observar os dados da amoxicilina. Contudo, as cidades com um baixo atendimento de esgoto podem ser consideradas prioritárias, tendo em vista que, para uma grande parte do esgoto gerado, esses medicamentos chegam diretamente aos ambientes aquáticos (Instituto Brasileiro de Geografia e Estatística, 2022).

Conclusão

A presente pesquisa demonstrou a viabilidade e eficácia da aplicação da linguagem de programação R como ferramenta para avaliação preliminar do risco ambiental causado por fármacos nos esgotos das capitais brasileiras. A metodologia aplicada, baseada na manipulação e visualização de grandes volumes de dados através de pacotes específicos do R, como o tidyverse e data.table, demonstrou ser uma alternativa eficiente e de baixo custo para complementar campanhas de monitoramento baseadas em coletas amostrais. O cálculo da Concentração Ambiental Prevista (CAP) possibilitou uma avaliação preliminar do risco potencial, permitindo a identificação de hotspots de contaminação nas diferentes capitais brasileiras. Estes resultados fornecem subsídios importantes para o direcionamento de políticas públicas de regulamentação específica para poluentes farmacêuticos no Brasil, bem como para o desenvolvimento de estratégias de tratamento de esgoto mais eficazes para a remoção destes compostos.

A existência de bancos de dados abertos, similares ao usado como base para este estudo, bem como registro para uma grande quantidade de medicamentos, possibilita a realização de análises preliminares para várias cidades de maneira simultânea e confiável.

Referências

ABRANTES, P. de M. *et al.* Avaliação da qualidade das prescrições de antimicrobianos dispensadas em unidades públicas de saúde de Belo Horizonte, Minas Gerais, Brasil, 2002. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 23, n. 1, p. 95-104, jan. 2007. Disponível em: <https://www.scielo.br/j/csp/a/KvXXQ6Bh63P7sM3XpStB3rc/?format=pdf&lang=pt>.

Acesso em: 15 mar. 2024.

Arnold J (2024). `_ggthemes`: Extra Themes, Scales and Geoms for 'ggplot2'_. R package version 5.1.0, <<https://CRAN.R-project.org/package=ggthemes>>.

Auguie B (2017). `_gridExtra`: Miscellaneous Functions for "Grid"Graphics_. R package version 2.3, <<https://CRAN.R-project.org/package=gridExtra>>.

Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, Schwendinger B, Krylov I (2025). `_data.table`: Extension of 'data.frame'_. R package version 1.17.0, <<https://CRAN.R-project.org/package=data.table>>.

BERQUÓ, L. S. *et al.* Utilização de antimicrobianos em uma população urbana. *Revista de Saúde Pública*, São Paulo, v. 38, n. 2, p. 554-560, 2004. Disponível em: <https://www.scielo.br/j/rsp/a/vKdG7TZxxQr4JjFW48PtV5v/?format=pdf&lang=pt>.

Acesso em: 15 mar. 2024.

CASTRO, P. B. N. e *et al.* Antibiotic consumption in developing countries defies global commitments: an overview on Brazilian growth in consumption. *Environmental Science and Pollution Research*, [s. l.], v. 27, p. 18515-18524, 2020.

EUROPEAN MEDICINES AGENCY. Committee for Medicinal Products for Human Use. *Guideline on the environmental risk assessment of medicinal products for human use*. London: EMEA, 2006. Disponível em: <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-environmental-risk-assessment>

Acesso em: 20 mar. 2024.

FC M, Davis T, ggplot2 authors (2025). `_ggpattern`: 'ggplot2' Pattern Geoms_. R package version 1.1.4, <<https://CRAN.R-project.org/package=ggpattern>>.

Firke S (2024). `_janitor`: Simple Tools for Examining and Cleaning Dirty Data_. R package version 2.2.1, <<https://CRAN.R-project.org/package=janitor>>.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016

HUTTNER, A. *et al.* Oral amoxicillin and amoxicillin-clavulanic acid: properties, indications and usage. *Clinical Microbiology and Infection*, Oxford, v. 26, p. 871-879, 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo demográfico*: Tabela 6805 – domicílio particulares permanentes ocupados, por tipo de esgotamento sanitário. Rio de Janeiro: IBGE, 2022. Disponível em: <https://sidra.ibge.gov.br/Tabela/6805>. Acesso em: 15 mar. 2024.

Mangiafico, Salvatore S. (2025). `rcompanion`: Functions to Support Extension Education Program Evaluation. version 2.5.0. Rutgers Cooperative Extension. New Brunswick, New Jersey. <https://CRAN.R-project.org/package=rcompanion>

Ooms J (2025). `_writexl`: Export Data Frames to Excel 'xlsx' Format_. R package version 1.5.2, <<https://CRAN.R-project.org/package=writexl>>.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, CIÊNCIA E CUL-

TURA. *Emerging Pollutants: Protecting Water Quality for the Health of People and the Environment*. [S. l.]: International Water Resources Association, [2023]. Disponível em: <https://iwraonlineconference.org/>. Acesso em: 15 mar. 2024.

Pedersen T, Robinson D (2024). *gganimate: A Grammar of Animated Graphics*. R package version 1.0.9, <https://CRAN.R-project.org/package=gganimate>

Petersen AH, Ekstrøm CT (2019). “dataMaid: Your Assistant for Documenting Supervised Data Quality Screening in R.” *Journal of Statistical Software*, *90*(6), 1-38. doi:10.18637/jss.v090.i06 <<https://doi.org/10.18637/jss.v090.i06>>.

QUINTILESIMS INSTITUTE. *Outlook for global medicines through 2021: balancing cost and value*. Parsippany, 2016. Disponível em: [https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/global-outlook-for-medicines-through-2021](https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/global-outlook-for-medicines-through-2021.pdf). Acesso em: 15 mar. 2024.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

Ryu C (2024). *dlookr: Tools for Data Diagnosis, Exploration, Transformation*. R package version 0.6.3, <<https://CRAN.R-project.org/package=dlookr>>.

SALES, P. de T. F. *et al.* Estudo da tratabilidade de efluente da indústria farmacêutica por meio dos fungos *Pycnoporus sanguineus*, *Schizophyllum commune* e fotocatalise. *Revista Eletrônica de Engenharia Civil*, Goiânia, v. 5, n. 1, p. 56-74, 2012. Disponível em: <https://revistas.ufg.br/reec/article/view/20257/12101>. Acesso em: 14 mar. 2024.

SILVA, R. F. da. *Avaliação de risco ambiental de fármacos*. 2019. Tese (Doutorado em Sistemas de Gestão Sustentáveis) – Escola de Engenharia, Universidade Federal Fluminense, Niterói, 2019. Disponível em: <https://app.uff.br/riuff/bitstream/handle/1/12776/Tese%20Ronaldo%20Ferreira%20da%20Silva.pdf?sequence=1>. Acesso em: 14 mar. 2024.

VAN BOECKEL, T. P. *et al.* Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *The Lancet Infectious Diseases*, [s. l.], v. 14, n. 8, p. 1-9, 2014.

WHITE, A. R. *et al.* Augmentin® (amoxicillin/clavulanate) in the treatment of community-acquired respiratory tract infection: a review of the continuing development of an innovative antimicrobial agent. *Journal of Antimicrobial Chemotherapy*, London, v. 53, p. i3-i20, 2004.

Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0, <<https://CRAN.R-project.org/package=forcats>>.

Wickham H (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1, <<https://CRAN.R-project.org/package=stringr>>.

Wickham H, Bryan J (2025). *readxl: Read Excel Files*. R package version 1.4.5, <<https://CRAN.R-project.org/package=readxl>>.

Wickham H, et. al. (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <<https://CRAN.R-project.org/package=dplyr>>.

Wickham H, Pedersen T, Seidel D (2023). *scales: Scale Functions for Visualization*. R package version 1.3.0, <<https://CRAN.R-project.org/package=scales>>.

Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, <<https://CRAN.R-project.org/package=tidyr>>.

Wilke C (2024). *gggridges: Ridgeline Plots in 'ggplot2'*. R package version 0.5.6, <<https://CRAN.R-project.org/package=gggridges>>.

William Revelle (2025). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.5.3, <<https://CRAN.R-project.org/package=psych>>.

Anexo

<https://github.com/simaourpia/AN-LISE-DE-DADOS-APLICADA-COMERCIALIZA-O-DE-DEZ-F-I>