



AVALIAÇÃO DA PREVALÊNCIA DE CÓLERA EM DADOS HOSPITALARES: UM ESTUDO COMPARATIVO ENTRE 2021 E 2022

GERLEY ADRIANO MIRANDA CRUZ; SAMUEL PIKHARDT MARTINS; ANA PAULA BEIRIGO BARBOSA; ANGÉLICA LIMA BRANDÃO SIMÕES

RESUMO

Introdução: No mês de setembro de 2023, a OMS (Organização Mundial da Saúde) emitiu um alerta para o aumento de notificações de cólera no mundo, uma vez que 44 países reportaram casos da doença (WHO, 2023). À luz desse panorama, no Brasil há uma evidente preocupação em relação às doenças transmitidas de forma hídrica, dado que a cólera — doença causada pela bactéria *Vibrio cholerae* — possui a água como seu principal meio de transmissão (SILVA, 2020). **Objetivo:** Avaliar se no Brasil houve uma diferença significativa para as internações e óbitos identificadas pelos registros de morbidade hospitalar do SUS (Sistema Único de Saúde), por local de internação, para os municípios que apresentaram registros da doença. **Materiais e Métodos:** Para a condução do estudo, foi aplicado o teste de Shapiro-Wilk para a verificação de um possível comportamento gaussiano para as variáveis “internações” e “óbitos” nos anos de 2021 e 2022. Após a conclusão do teste de normalidade, o teste U de Mann-Whitney foi utilizado para avaliar a existência de diferenças consideráveis para as variáveis previamente selecionadas. **Resultados:** Durante a fase de coleta foram identificadas 164 municípios com casos de cólera no ano de 2021. Para o ano posterior, 196 municípios apresentaram notificações da doença. Em ambos os anos, para todas as variáveis de interesse, não foi observado uma distribuição normal. Tal comportamento permitiu o uso do teste U de Mann-Whitney, no qual foi obtido, para o parâmetro “internações”, uma estatística de teste de 15503,0, enquanto o valor-p foi de 0,52. Por outro lado, para a variável “óbitos”, a estatística de teste foi de 16740,5 enquanto o que valor-p foi de 0,202. Por conseguinte, não foram observadas diferenças significativas para os anos selecionados. **Conclusões:** O estudo possibilitou verificar a inexistência de um desvio de normalidade observado entre os anos de 2021 e 2022. Essa análise, contudo, não exige a necessidade de estratégias públicas para o melhor atendimento e oferta de um saneamento básico à população.

Palavras-chave: Cólera, *Vibrio cholerae*, Análise Epidemiológica, Significância Estatística.

1. INTRODUÇÃO

O alerta para o aumento de casos de cólera no mundo, pela própria OMS (Organização Mundial da Saúde), no mês de setembro de 2023, reacendeu uma preocupação quanto à iminência de uma nova emergência sanitária de magnitude internacional. O crescimento

aproximado de 25% para o ano de 2022, em relação ao ano anterior, segundo dados da própria instituição, evidencia não somente uma necessidade de políticas públicas de conscientização concernente à prevenção de contágio do patógeno *Vibrio cholerae*, mas também a urgência de um entendimento analítico desse comportamento epidemiológico em alarde.

A cólera é caracterizada como uma fisiopatologia infecciosa acometida pela enterotoxina secretada pelo agente bacteriano gram-negativo *Vibrio cholerae* — bacilo em forma de bastão, com respiração aeróbica ao mesmo tempo que dotado de mobilidade devido à presença de um flagelo polar (LOPES, 1992). O quadro sintomatológico, por sua vez, é comumente dividido em três períodos: início ou da diarreia premonitória, cólera propriamente dita e o estágio de reação. São nesses períodos em que são diferenciados os sinais sintomas mais evidentes, tais como a diarreia com fezes líquidas acompanhadas de uma particular coloração na fase inicial, vômitos frequentes e câimbras gástricas durante o estágio de cólera e o desaparecimento dos sintomas de algidez e normalização da circulação e das funções orgânicas ao longo da etapa de reação (QUEIRÓZ, 1911).

Já foram registradas, ao todo, desde o século XIX, um total de sete pandemias de cólera – tendo ocorrido pela primeira vez na porção oriental da Índia, em Calcutá, no ano de 1817 (LAFUENTE). No Brasil, as precárias condições de saneamento básico foram constantemente caracterizadas como um dos principais fatores de agravamento para tais eventualidades (SILVA, 2020). No que tange às diferentes formas de contaminação, em seus mais variados modos de consumo – seja para ingestão, banho, preparo e irrigação de alimentos – a água é compreendida como o principal meio de propagação (LOPES, 1992).

Nesse sentido, a correlação entre um saneamento de qualidade junto à diminuição de doenças de veiculação hídrica é imperativa quando se trata de políticas de prevenção (QUEIRÓZ, 1911).

A busca por uma maior percepção do quadro atual de cólera no Brasil ajuda a compreender como o país situa-se diante de constantes surtos em territórios como o Afeganistão, Camarões, República Democrática do Congo, Malawi, Nigéria, Somália, Síria e República Árabe (WHO, 2023, p. 431). De modo geral, em 2022, 80 países forneceram dados sobre a cólera para a OMS. Entre esses países, 44 relataram um total de 472.697 casos e 2.349 óbitos, resultando em uma taxa de letalidade (CFR) de 0,5%. Sob essa análise, o Brasil não é mencionado de forma proeminente no relatório intitulado “Weekly Epidemiological Record, 2023, vol. 98, 38”. Contudo, para averiguar a evolução dos casos de infecção por *Vibrio cholerae* no país, a averiguação de dados acompanhado de testes de significância estatística, tanto em relação a configuração de normalidade dos registros, quanto às diferenças significativas para os anos destacados pela OMS, é uma ferramenta crucial para o maior conhecimento desse cenário.

2. MATERIAIS E MÉTODOS

A seleção e coleta dos dados foram feitas por meio de pesquisas realizadas na plataforma TABNET – ferramenta de tabulação desenvolvida pelo DATASUS. A consulta levou em consideração as morbidades hospitalares registradas pelo SUS (Sistema Único de Saúde), por local de internação para os municípios do Brasil que apresentaram registros para os anos de 2021 e 2022. Os conteúdos selecionados para posterior análise foram “internações” e “óbitos”. Apenas a opção cólera foi marcada como parte do filtro para o parâmetro “Lista Morb CID-10”. Os dados foram exportados separadamente, isto é, para cada um dos anos, como arquivo tipificados no formato CSV (acrônimo para Comma-separated values).

A linguagem de programação utilizada para codificação das etapas do presente estudo foi a linguagem Python, em sua versão 3.11. As bibliotecas Pandas, Matplotlib, Seaborn e Scipy foram utilizadas para coleta e serialização dos dados, visualização, refinamento gráfico

e realização de testes estatísticos, respectivamente. Todo o processo de codificação foi realizado em um ambiente virtual de desenvolvimento conhecido como Github Codespace - recurso mantido pela Microsoft por meio do sistema de hospedagem e versionamento de repositórios Github.

Após a coleta, mediante a uma análise geral dos dados, em especial se tratando da quantidade de registros para cada um dos parâmetros selecionados (internações e óbitos), os dados não informados, classificados como nulos, foram substituídos por zero. Foi avaliado uma possibilidade de substituição utilizando abordagens de machine learning, seja por meio de regressões, KNN (K-nearest neighbors – K-ésimo Vizinho mais Próximo) ou até mesmo técnicas de clusterização – avaliando a segmentação de possíveis grupos (clusters) com dados homogêneos. Contudo, dado a abundância de valores não identificados para a coluna “óbitos”, e tendo como base a regularidade quantitativa para os dados selecionados, foi optado por entender como a ausência de casos de cólera para tais registros.

A verificação da normalidade dos dados, também conhecida como distribuição gaussiana, foi feita por meio do teste de Shapiro-Wilk (GHASEMI et al, 2012). A técnica tem como finalidade verificar a hipótese nula de que uma amostra $x_1, x_2, \dots, x_n$, seccionada de uma população, possui um comportamento gaussiano. Desse modo, pressupõe que a hipótese alternativa compreende um comportamento contrário ao descrito pela hipótese nula. O valor de estatística W , – coeficiente para posterior validação, usada para medir o grau de desvio dos dados da amostra em relação a uma distribuição normal –, pode ser calculado da seguinte forma:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

A equação logo acima é utilizada para obtenção do valor de estatística W . Onde, $x_{(i)}$ representa cada uma das observações (depois de ordenadas da menor para a maior). Cada observação se refere aos dados do conjunto de “internações” ou “óbitos”. O parâmetro x é caracterizado como sendo a média aritmética dos registros para o conjunto de interesse. Os valores de a_i referem-se aos coeficientes previamente tabulados.

O valor-p (abreviação para "p-value" em inglês), coeficiente usado para indicar a concreta probabilidade de se encontrar uma diferença significativa em relação à que foi observada sob a hipótese nula (FERREIRA et al, 2015) e estabelecido como 0,05. A validação da normalidade dos dados ao comparar com o valor-p, previamente instituído, foi crucial para a escolha do teste de avaliação da variação expressiva entre os anos delimitados. Essa decisão é explicada pela dependência do comportamento dos dados em relação a sua simetria em torno de um valor central.

Em virtude do comportamento não gaussiano evidenciado pelo teste de Shapiro-Wilk, o teste de hipótese usado para verificar se houve uma discrepância considerável para as variáveis “internações” e “óbitos” nos municípios do Brasil para os anos de 2021 e 2022, foi o Teste U de Mann-Whitney. O teste consiste em avaliar, em dois grupos não pareados, possivelmente não normais, de forma não paramétrica, a igualdade das medianas (MCKNIGHT e NAJAB, 2010). Para avaliação da significância estatística das diferenças desejadas, assim como realizado no exame anterior de normalidade, foi utilizado o valor-p de 0,05.

O sequenciamento da codificação das etapas do estudo foi realizado utilizando a ferramenta Jupyter Notebook – uma ferramenta de software livre comumente usada para

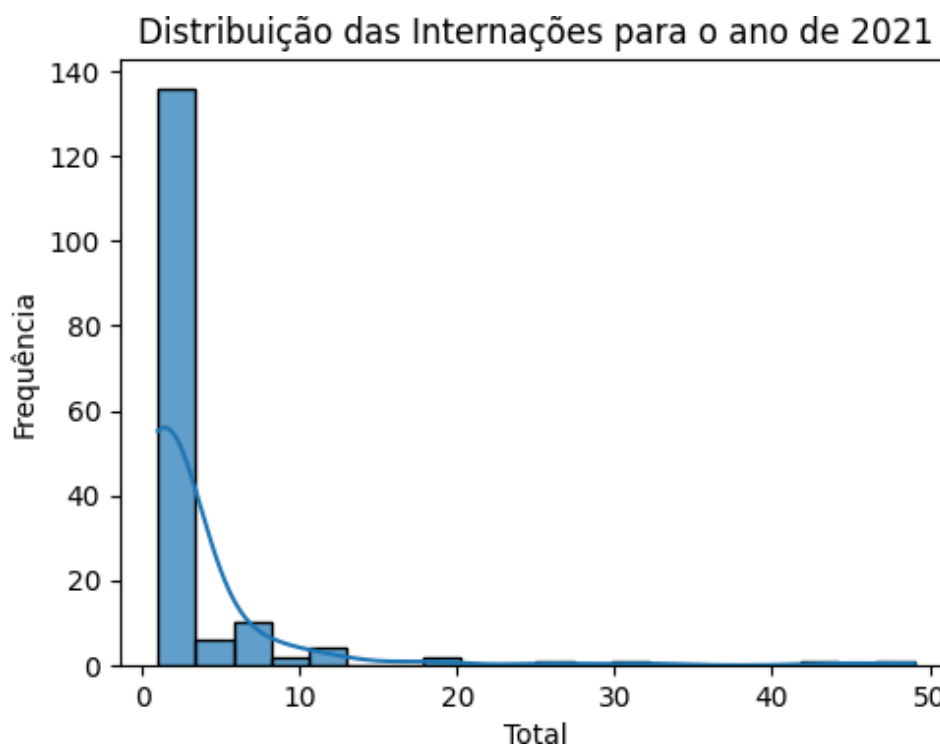
análise de dados, visto o seu recurso de interatividade em blocos programáveis.

3. RESULTADOS E DISCUSSÃO

Ao todo, com base nos critérios de seleção delimitados, foram selecionados 164 municípios com casos de cólera no Brasil no ano de 2021. Enquanto no ano de 2022, foram encontrados 196 municípios – um aumento aproximado de 19% em relação ao ano anterior, no que se refere exclusivamente a quantidade de registros. A média de internações para os dois períodos foi respectivamente 3,274 e 3,127. Para o ano de 2021 foram identificados 21 óbitos, por outro lado, para o ano de 2022, o valor obtido foi de 17.

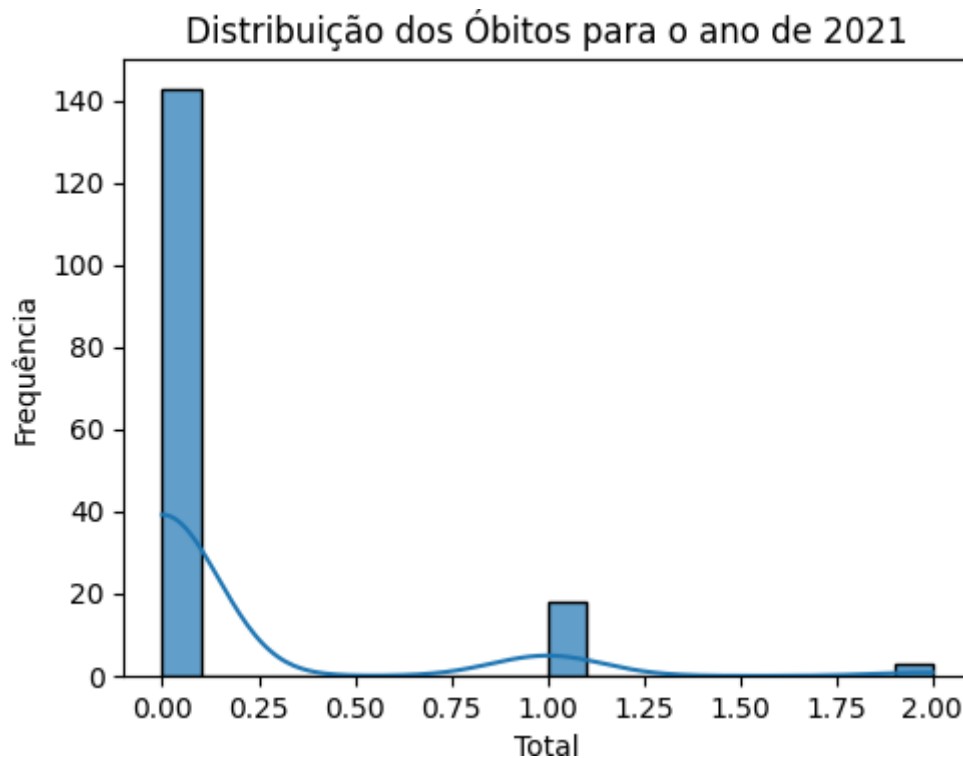
Para o teste de Shapiro-Wilk, em relação ao ano de 2021, no que tange a variável “internações”, a estatística de teste encontrada foi de aproximadamente 0,388. O valor-p foi de $2,31 \times 10^{-23}$. Para o parâmetro “óbitos”, a estatística de teste e o valor-p foram respectivamente valores aproximados de 0,4 e $3,73 \times 10^{-23}$. Assim, denota-se que em ambas as colunas não há o seguimento de uma distribuição normal.

Figura 1 – Distribuição das internações para o ano de 2021 com base nos dados selecionados para a morbidade hospitalar do SUS para os municípios do país que notificaram casos de cólera



Fonte: Construção da autoria

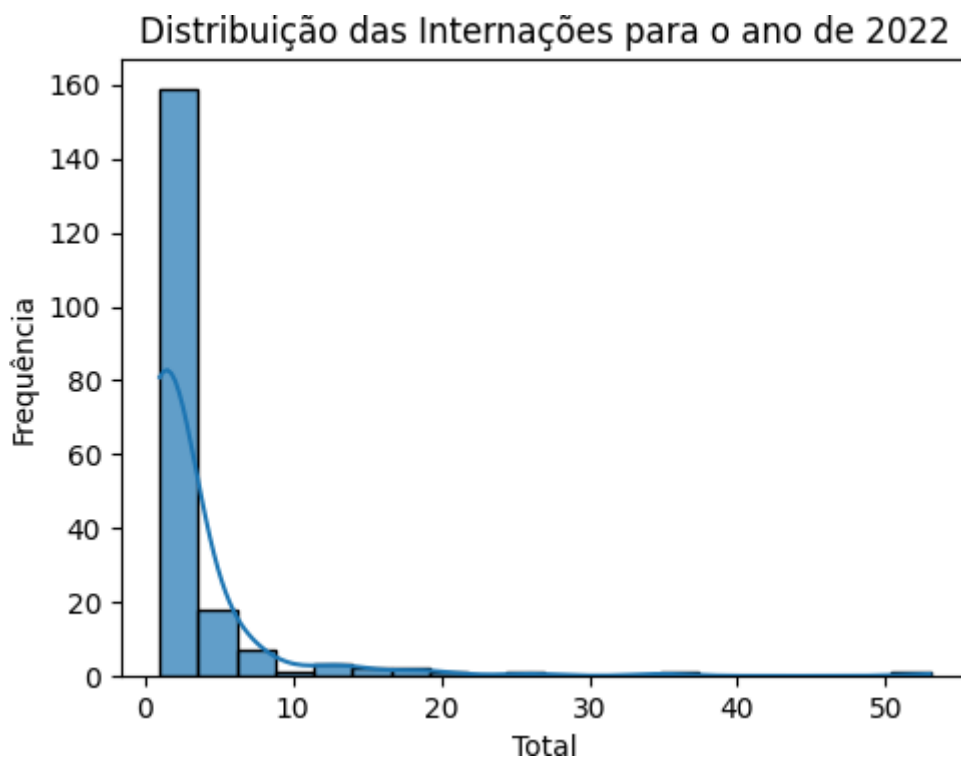
Figura 2 –Distribuição dos óbitos para o ano de 2021 com base nos dados selecionados para a morbidade hospitalar do SUS para os municípios do país que notificaram casos de cólera.



Fonte: Construção da autoria

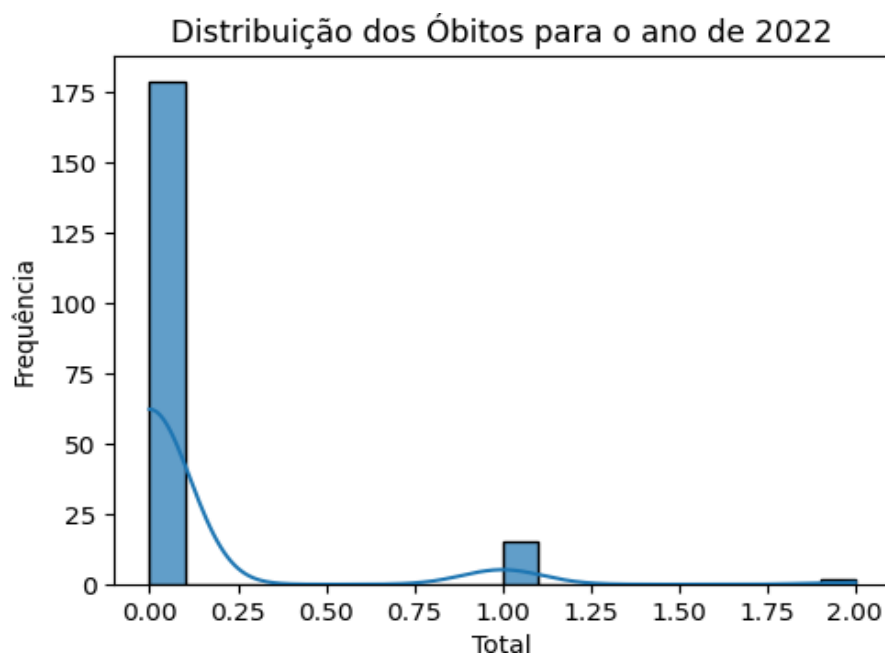
Em relação ao teste de Shapiro-Wilk para os dados de 2022, foram obtidos para as variáveis “internações” e “óbitos” os valores de estatística de teste 0,407 e 0,318. Para o valor-p, em relação a essas duas colunas, os valores 5.24×10^{-25} e 1.796×10^{-26} foram encontrados. Assim como observado para os dados do ano de 2021, tais saídas também denotam um comportamento não gaussiano para os dados.

Figura 3 – Distribuição das internações para o ano de 2022 com base nos dados selecionados para a morbidade hospitalar do SUS para os municípios do país que notificaram casos de cólera.



Fonte: Construção da autoria

Figura 4 – Distribuição dos óbitos para o ano de 2022 com base nos dados selecionados para a morbidade hospitalar do SUS para os municípios do país que notificaram casos de cólera.



Fonte: Construção da autoria

Sabendo que os dados não seguem o padrão de normalidade esperado, a escolha de um teste não paramétrico foi necessária para avaliação de uma possível diferença significativa entre os dois anos em questão. Conforme essa perspectiva, para a variável “internações”, a estatística de teste encontrada foi de, 15503,0, enquanto o Valor-p foi de 0,520. Já para o parâmetro “óbitos”, a estatística de teste em evidência foi de 16740,5 enquanto o Valor-p foi de 0,202. Dessa maneira, percebe-se que, ao assumir o Valor-p de 0,05 como critério para validação da significância estatística, em relação aos casos de cólera para os municípios do país em que houve registros notificados, não houve uma mudança expressiva entre os anos de 2021 e 2022 para os dois parâmetros selecionados para o estudo.

Figura 5 – Análise da diferença entre os dois valores-p (internações e óbitos) nos anos de 2021 e 2022 para o teste U de Mann- Whitney



Fonte: Construção da autoria

4. CONCLUSÃO

Portanto, conclui-se que, apesar do alerta para o aumento de casos de cólera no mundo, pela própria OMS, no Brasil, nos anos comparados pela organização, no que se refere às morbidades hospitalares registradas pelo SUS, não houve uma diferença considerável, entre os dois anos, que qualifique tal cenário como de iminente calamidade. Entretanto, tal análise não suprime a urgência de políticas públicas voltadas para a democratização do saneamento básico no país.

Tal premissa dialoga diretamente com o terceiro e o sexto objetivo da Agenda 2030 da ONU: a promoção da saúde e do bem-estar, além do acesso à água potável e saneamento (UNITED NATIONS GENERAL ASSEMBLY, p. 16 e p. 18). Logo, com base nos métodos e ferramentas utilizados, a coleta evidenciou casos de cólera em alguns municípios do país, sem discrepâncias comparado ao mesmo período -2021 e 2022- no internacional .

Torna-se convidativo o prosseguimento desse estudo para uma abrangência de novas correlações que permitam uma análise mais detalhada das características regionais que caracterizam os casos de cólera nessas localidades.

REFERÊNCIAS

FERREIRA, Juliana Carvalho; PATINO, Cecilia Maria. O que realmente significa o valor-p?. **Jornal Brasileiro de Pneumologia**, v. 41, p. 485-485, 2015.

GHASEMI, Asghar; ZAHEDIASL, Saleh. "Normality tests for statistical analysis: a guide for non- statisticians." **International Journal of Endocrinology and Metabolism**, v. 10, n. 2, p. 486-9, 2012. DOI: 10.5812/ijem.3505.

LAFUENTE, Sarah et al. El cólera. **Enfermedades Emergentes**, v. 8, n. 1, p. 10-15, 2006.
LOPES, Hélio Vasconcelos; AYUB, Munir Akar. Cólera. **Arquivos Médicos do ABC**, v. 15, n. 1, 1992. MCKNIGHT, Patrick E.; NAJAB, Julius. Mann?Whitney U Test. **The Corsini Encyclopedia of Psychology**, p. 1-1, 2010.

QUEIRÓZ, Alexandre. A Cólera. 1911. SILVA, Everaldo de Santana; OLIVEIRA, Deloar Duda de; LOPES, Amanda Pontes. Acesso ao Saneamento básico e Incidência de Cólera: uma análise quantitativa entre 2010 e 2015. **Saúde em debate**, v. 43, p. 121-136, 2020.

UNITED NATIONS GENERAL ASSEMBLY. **Transforming our world: the 2030 Agenda for Sustainable Development**, 21 de Outubro de 2015, A/RES/70/1. Disponível em: <https://www.refworld.org/docid/57b6e3e44.html>. Acesso em: 14 de outubro de 2023.

WORLD HEALTH ORGANIZATION. Weekly Epidemiological Record, 2023, vol. 98, n. 38 [full issue]. **Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire**, v. 98, n. 38, p. 431 - 452, 2023-09-22.