Correlação e sensibilidade do Google Trends para surtos de dengue e febre amarela no estado de São Paulo

Google Trends correlation and sensitivity for outbreaks of dengue and yellow fever in the state of São Paulo

Vitor Ulisses Monnaka¹, Carlos Augusto Cardim de Oliveira¹

¹ Faculdade Israelita de Ciências da Saúde Albert Einstein, Hospital Israelita Albert Einstein, São Paulo, SP, Brasil.

DOI: 10.31744/einstein_journal/2021AO5969

■ RESUMO

Objetivo: Avaliar a acurácia do Google Trends para vigilância epidemiológica de dengue e febre amarela e comparar a incidência dessas doenças com a popularidade de seus termos no estado de São Paulo. Métodos: Coorte retrospectiva. Os resultados da pesquisa Google Trends foram comparados com a incidência real de doenças, obtida do Centro de Vigilância Epidemiológica “Prof. Alexandre Vranjac”, do estado de São Paulo, nos períodos entre 2017 e 2019. A correlação foi calculada pelo coeficiente de Pearson e pela função de correlação cruzada. A acurácia foi analisada por valores de sensibilidade e especificidade. Resultados: Houve correlação estatisticamente significante entre as variáveis estudadas para ambas as doenças, com coeficiente de Pearson de 0,91 para dengue e 0,86 para febre amarela. Foi identificada correlação com até 4 semanas de antecipação para séries temporais. A sensibilidade foi de 87% e 90% e a especificidade de 69% e 78% para dengue e febre amarela, respectivamente. Conclusão: A incidência de dengue e febre amarela no estado de São Paulo apresentou forte correlação com a popularidade de seus termos medidos pelo Google Trends em períodos semanais. A ferramenta Google Trends forneceu alerta precoce, com alta sensibilidade, para a detecção de surtos dessas doenças.

Descritores: Doenças transmissíveis; Monitoramento epidemiológico; Vigilância da população; Ferramenta de busca; Tecnologia da informação; Previsões; Febre amarela; Dengue

■ ABSTRACT

Objective: To assess Google Trends accuracy for epidemiological surveillance of dengue and yellow fever, and to compare the incidence of these diseases with the popularity of its terms in the state of São Paulo. Methods: Retrospective cohort. Google Trends survey results were compared to the actual incidence of diseases, obtained from Centro de Vigilância Epidemiológica “Prof. Alexandre Vranjac”, in São Paulo, Brazil, in periods between 2017 and 2019. The correlation was calculated by Pearson’s coefficient and cross-correlation function. The accuracy was analyzed by sensitivity and specificity values. Results: There was a statistically significant correlation between the variables studied for both diseases, Pearson coefficient of 0.91 for dengue and 0.86 for yellow fever. Correlation with up to 4 weeks of anticipation for time series was identified. Sensitivity was 87% and 90%, and specificity 69% and 78% for dengue and yellow fever, respectively. Conclusion: The incidence of dengue and yellow fever in the State of São Paulo showed a strong correlation with the popularity of its terms measured by Google Trends in weekly periods. Google Trends tool provided early warning, with high sensitivity, for the detection of outbreaks of these diseases.

Keywords: Communicable diseases; Epidemiological monitoring; Population surveillance; Search engine; Information technology; Forecasting; Yellow fever; Dengue
INTRODUÇÃO

As doenças transmissíveis são uma ameaça à saúde dos indivíduos, principalmente nos países em desenvolvimento. No Brasil, a dengue e a febre amarela são infeções de grande impacto na saúde da população. A identificação precoce de surtos de doenças transmissíveis aumenta a possibilidade de controle da disseminação, com uso provável em intervenções para prevenção, isolamento e tratamento. A dengue é uma infecção aguda, com altas taxas de mortalidade, transmitida predominantemente pelo mosquito *Aedes aegypti*, cujos casos são classificados como sem sinais de alerta, com sinais de alerta e graves, de acordo com a Organização Mundial da Saúde (OMS). É a principal infecção por arbovírus no mundo, e o Brasil é o país com o maior número de casos nesse século. A febre amarela, uma infecção aguda, febril, hemorrágica e não contagiosa, responde por alta mortalidade na população da América do Sul e da África, desde o século 17. Ela se manifesta em dois padrões de transmissão epidemiológica distintos: selvagem e urbano – ambos pelo *Aedes aegypti*. De acordo com o Boletim Epidemiológico volume 51, publicado pelo Ministério da Saúde do Brasil, 714.164 casos prováveis de dengue foram identificados, e 298 mortes por dengue foram confirmadas de janeiro a maio de 2020. Entre julho de 2019 e maio de 2020, 812 casos de febre amarela foram registrados no país, dos quais 324 ocorreram no estado de São Paulo.

Devido ao impacto na saúde da população brasileira, a vigilância efetiva de casos de dengue e febre amarela é extremamente importante para o controle de epidemias. No estado de São Paulo, o Centro de Vigilância Epidemiológica (CVE) “Prof. Alexandre Vranjac”, instituição subordinada à estrutura da Coordenadoria de Controle de Doenças (CDC), é responsável pela divulgação de relatórios periódicos sobre o status dessas doenças, por semana epidemiológica. No entanto, a presença de uma estrutura de vigilância efetiva não é homogênea em todos os estados do Brasil, e a subnotificação é um possível fator de falha para acurácia do número de casos.

Considerando-se os obstáculos presentes na vigilância epidemiológica de doenças transmissíveis, já foi sugerido o uso de ferramentas on-line como métodos complementares para obtenção de informações, que sinalizem possíveis surtos de doenças. Um estudo com dados brasileiros usando a ferramenta Twitter demonstrou associação entre *tweets* e a dengue e levantou a possibilidade de essa ferramenta ser usada para estimar o número de casos semanalmente. O Plano de Contingência Nacional para Epidemias de Dengue, elaborado pelo Ministério da Saúde, também menciona o uso de tendências relativas de boatos no Twitter como indicadores para ações de resposta específicas. Nesse sentido, o Google Trends, ferramenta que analisa a popularidade de termos de busca no Google, em períodos e locais determinados, poderia ser útil para a vigilância de casos de dengue e febre amarela.

No Google Trends, a tendência de um determinado termo é exibida em uma escala de zero a cem, em que cem representa o maior volume de buscas para o termo em questão, em determinado local e período. Os resultados representam um valor relativo, que reflete o número de buscas realizadas para um termo específico, comparado ao total de número de buscas realizadas. Recentemente, as tentativas de usar esse instrumento para questões relativas à saúde têm sido mais frequentes. Estudos anteriores já analisaram a capacidade do Google Trends de prever epidemias de gripe na América Latina e casos confirmados de Zika e demonstraram correlação entre as tendências de termos e os casos de dengue no mundo.

Sendo assim, considerando-se as evidências já publicadas, é importante avaliar a utilidade dessa plataforma para vigilância epidemiológica da dengue e da febre amarela.

OBJETIVO

Avaliar a acurácia do Google Trends para vigilância epidemiológica da dengue e da febre amarela e comparar a incidência dessas doenças com a popularidade de seus termos no estado de São Paulo.

MÉTODOS

O projeto foi realizado na Faculdade Israelita de Ciências da Saúde Albert Einstein (FICSAE - HIAE), de agosto de 2018 a agosto de 2019. Este projeto não precisou de aprovação do Comitê de Ética em pesquisa, porque usou somente dados de domínio público, sem envolver seres humanos. O desenho usado foi o de coorte retrospectiva. Os dados do Google Trends foram obtidos no plataforma on-line https://trends.google.com/trends/, que informa as tendências relacionadas à frequência de buscas dos respectivos termos no Google, em uma escala de zero a cem, na qual cem representa o maior volume de buscas pelo termo, em um determinado local e período. Os resultados representados como “<1” em termos de tendência foram arredondados para o valor de umx, com o objetivo de padronizar quantitativamente as informações, permitindo a análise estatística. Informações sobre a febre amarela foram obtidas com base no termo “febre amarela” no estado de São Paulo, por
Correlação e sensibilidade do Google Trends para surtos de dengue e febre amarela no estado de São Paulo

semana epidemiológica, de 1º de janeiro de 2017 a 19 de maio de 2018, em um total de 70 semanas, sendo 50 de 2017 e 20 de 2018. As semanas epidemiológicas 24 e 26, de 2017, foram excluídas da análise devido à falta de publicação de boletins epidemiológicos para esses períodos. Informações sobre a dengue foram obtidas com base no termo “dengue” no estado de São Paulo, por semana epidemiológica, de 31 de dezembro de 2017 a 30 de março de 2019, em um total de 65 semanas, sendo 52 de 2018 e 13 de 2019. A incidência de febre amarela e dengue no estado de São Paulo foi obtida no boletim epidemiológico divulgado pelo CVE da Secretária de Saúde do Estado de São Paulo, que lista o número total de casos nos períodos correspondentes.

Análise estatística
A associação entre as medições quantitativas dos métodos foi avaliada usando-se o coeficiente de correlação de Pearson e a análise de séries históricas.\(^{(21)}\) A função de correlação cruzada permite avaliar a dependência temporal entre as séries de variáveis por meio dos valores de \(\text{lag}\), que expressam o grau e a direção da associação. Um \(\text{lag}\) de -2 para um determinado coeficiente indica que os dados do Google Trends corresponderam a 2 semanas a menos que os registros do CVE. Isso quer dizer que a correlação do aumento nas tendências é notada 2 semanas antes do registro dos casos. Foram realizadas análises estatísticas com o programa RStudio,\(^{(22)}\) e o nível de significância considerado foi 0,05.

A acurácia diagnóstica da ferramenta para detecção de epidemias no estado de São Paulo foi avaliada pela classificação das semanas epidemiológicas, quanto à presença ou à ausência de epidemias, e tendências nos termos do Google Trends, pela presença ou ausência de sinais de alerta. Com base nos dados de CVE, considerados o padrão-ouro, estabelecemos como limiares epidemiológicos os números de casos de, no mínimo, três e 500, para febre amarela e dengue, respectivamente. Consideramos como sinal de alerta valores de tendências de, no mínimo, quatro e cinco, para febre amarela e dengue, respectivamente. A comparação entre esses dados permitiu o cálculo dos valores de sensibilidade e especificidade.

II RESULTADOS
As figuras 1 e 2 representam a incidência de doenças e seus termos no Google Trends para dengue e febre amarela, respectivamente. Na figura 1, a epidemia de dengue foi determinada a partir da semana 50, quando ocorreram 798 casos. Na figura 2, há duas epidemias de febre amarela, uma começando na semana 11 e outra na semana 50, ambas com três casos.

As figuras 3 e 4 representam gráficos de dispersão entre a incidência da doença e os termos de busca no Google Trends para dengue e febre amarela, respectivamente. Houve correlação estatisticamente significativa (\(p<0,0001\)) para ambas as doenças. No caso da dengue, o coeficiente de Pearson foi 0,91; para a febre amarela, o coeficiente foi 0,86. Os resultados dessas análises são mostrados na tabela 1. A análise de correlação cruzada mostrou correlação estatisticamente significativa para até 4 semanas de deslocamento entre as séries históricas, conforme mostrado na tabela 2.
Para avaliar a acurácia do Google Trends para detecção de epidemias, os valores de positivos-verdadeiros, falsos-positivos, negativos-verdadeiros e falsos-negativos foram calculados, conforme mostrado na tabela 3. A partir desses dados, foram calculadas sensibilidade de 87% e especificidade de 69% para dengue, e sensibilidade de 90% e especificidade de 78%, para febre amarela.

### DISCUSSÃO

O Google Trends permite a avaliação do comportamento humano e a previsão de problemas de saúde, tendo sido demonstrado que a sazonalidade observada nas buscas on-line está relacionada a casos das doenças estudadas.

Métodos estatísticos e abordagens para esse tipo de análise já foram descritos em revisões sistemáticas.

O presente estudo demonstrou que os dados obtidos com a ferramenta tiveram forte correlação com a incidência de febre amarela no estado de São Paulo, na avaliação em intervalos semanais. Essa alta correlação entre os casos de dengue e o Google Trends já foi identificada na Indonésia, nas Filipinas e na Índia. Usando dados do Brasil, Yang et al. compararam levantamentos de casos de dengue fornecidos pelo Ministério da Saúde, mensalmente, de janeiro de 2001 a dezembro de 2012, e encontraram correlação de 0,971, semelhante ao resultado da presente pesquisa. No entanto, este estudo é o primeiro a demonstrar essa correlação com dados brasileiros de base semanal. Essa abordagem possibilitou avaliação de séries históricas já realizadas em outros países e demonstrou correlação moderada antes que a epidemia ocorresse, com até 4 semanas de diferença para dengue e 3 semanas para febre amarela. Isso indica que a ferramenta tem a habilidade de fornecer um alerta precoce, permitindo que as autoridades tomem medidas para prever a propagação dessas doenças.

Outros estudos avaliaram o Google Trends usando dados brasileiros de outras doenças. Um estudo avaliou a capacidade de previsão de epidemias de gripe na América Latina, comparando-se a proporção de casos na plataforma FluNet, entre janeiro de 2011 e dezembro de 2014, com dados do Google Trends, e os coeficientes de correlação de Pearson estiveram entre 0,48, em 2012, e 0,61, em 2014, no Brasil. Este artigo encontrou imprecisão considerável do Google Trends comparada a FluNet, muito provavelmente devido ao acesso limitado à Internet em algumas regiões. Destacou também as limitações da FluNet devido à dimensão geográfica do Brasil, bem como por sua diversidade ecológica e demográfica. Outro estudo analisou a capacidade de previsão de casos confirmados de Zika no Brasil e demonstrou que o Google Trends conseguiu prever a epidemia com 1 semana de antecipação.

Diferente do estudo de Marques-Toledo et al., que desenvolveram um modelo para prever o número de casos de dengue com base em dados da rede social Twitter, este estudo propôe um uso diferente das ferramentas on-line. Acredita-se que a maior importância desses instrumentos está na identificação da ocorrência de epidemias, e não necessariamente na previsão do
número de casos. Portanto, este estudo é o primeiro a analisar a acurácia do Google Trends na identificação de surtos de doenças. Foi observada alta sensibilidade para febre amarela (0,90) e dengue (0,87), indicando utilidade prática da ferramenta, principalmente porque, para fins de previsão, a sensibilidade é mais útil que a especificidade, já que a primeira indica baixa probabilidade de falsos-negativos e de casos perdidos. Portanto, a ferramenta pode ser muito útil, principalmente em estados com sistemas de vigilância epidemiológica menos eficazes, como uma análise complementar aos métodos disponíveis.

Outra utilidade relevante do Google Trends relativa à saúde poderia ser a avaliação de doenças que não estejam no escopo das instituições de vigilância epidemiológica. A ferramenta também poderia ser consideravelmente útil para o monitoramento de reações adversas incomuns e novos efeitos benéficos de medicamentos, bem como para avaliar a minimização de doses efetivas (estudos de Fase IV), após a inserção de medicamentos no mercado.

Limitações
Existem algumas limitações nos métodos de análise empregados neste estudo, para verificar a incidência da doença como causa do aumento das tendências de termos relacionados, devido à existência de outros mecanismos de associação entre duas variáveis, como acaso ou confusão. Como as tendências do Google Trends são determinadas pelos interesses dos usuários da Internet, elas podem produzir uma correlação aleatória, não necessariamente decorrente da incidência das doenças. No entanto, o maior volume de pesquisas reduz significativamente a probabilidade de erros, devido ao fator acaso. Além disso, pode haver outras variáveis responsáveis por induzir uma confusão positiva, como campanhas de conscientização conduzidas no momento de pico da incidência da doença ou menções na mídia, o que aumenta o interesse público e, consequentemente, a frequência de buscas.

Enfatiza-se também que, como o uso da ferramenta Google Trends depende do acesso à Internet, regiões menos favorecidas apresentam taxas de busca mais baixas, mesmo com alta incidência de determinada doença, o que limita a aplicabilidade deste instrumento.

CONCLUSÃO
O estudo demonstrou correlação significativa entre os dados gerados pela ferramenta Google Trends e a incidência de dengue e febre amarela no estado de São Paulo na avaliação em intervalos semanais. A maior frequência de buscas serviu como alerta precoce para surtos dessas doenças e demonstrou alta sensibilidade para a detecção de epidemias. São necessárias mais pesquisas para confirmar estes achados para outras doenças e localidades, mas as conclusões apontam para a possibilidade de se usar essa ferramenta como método simples e barato de vigilância epidemiológica.

CONTRIBUIÇÃO DOS AUTORES
Vitor Ulisses Monnaka: conceitualização, investigação, metodologia, coleta de dados, análise e escrita do manuscrito original. Carlos Augusto Cardim de Oliveira: investigação, metodologia, supervisão, revisão e edição do manuscrito.

INFORMAÇÃO DOS AUTORES
Monnaka VU: http://orcid.org/0000-0001-7313-0437
Oliveira CA: http://orcid.org/0000-0001-9764-0669

REFERÊNCIAS
1. Boutayeb A. The double burden of communicable and non-communicable diseases in developing countries. Trans R Soc Trop Med Hyg. 2006;100(3):191-9. Review.
2. Waldman EA, Sato AP. Path of infectious diseases in Brazil in the last 50 years: an ongoing challenge. Rev Saude Publica. 2016;22;50:68. Review.
3. Ellwanger JH, Kaminski VL, Chies JA. Emerging infectious disease prevention: Where should we invest our resources and efforts? J Infect Public Health. 2019;12(3):313-6. Review.
4. Mukhopadhyay S, Kuhn RJ, Rossmann MG. A structural perspective of the flavivirus life cycle. Nat Rev Microbiol. 2005;3(1):13-22. Review.
5. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, et al. Modeling tools for dengue risk mapping - a systematic review. Int J Health Geogr. 2014;13:50. Review.
6. Gibbons RV, Vaughn DW. Dengue: an escalating problem. BMJ. 2002;324(7353):1563-6. Review.
7. Teixeira MG, Costa MC, Barreto F, Barreto ML. Dengue: twenty-five years since reemergence in Brazil. Cad Saude Publica. 2009;25 Suppl 1:S7-18. Review.
8. Antonio FJ, Itami AS, de Picoli S, Teixeira JJ, Mendes RD. Spatial patterns of dengue cases in Brazil. PLoS One. 2017;12(7):e0180715.
9. Monath TP. Yellow fever: an update. Lancet Infect Dis. 2001;1(1):11-20. Review.
10. Reiter P. Yellow fever and dengue: a threat to Europe? Euro Surveill. 2010;15(10):19509.
11. Saad LD, Barata RB. Yellow fever outbreaks in São Paulo State, Brazil, 2000-2010. Epidemiol Serv Saude. 2016;25(3):531-40.
12. Marques-Toledo CA, Degener CM, Vinhal L, Coelho G, Meira W, Cedeño CT, et al. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. PLoS Negl Trop Dis. 2017;11(7):e0005729.
13. Google Trends. Califórnia: Google LLC; s.d. [cited 2019 Feb 15]. Available from: https://trends.google.com/trends/
14. Google LLC. Califórnia: Google LLC; s.d. [cited 2019 Feb 15]. Available from: https://www.google.com/
15. Pollett S, Boscardin WJ, Aziz-Baungartner E, Tinoco YO, Soto G, Romero C, et al. Evaluating Google Flu Trends in Latin America: important lessons for the next phase of digital disease detection. Clin Infect Dis. 2017;64(1):34-41.

16. Morsy S, Dang TN, Kamel MG, Zayan AH, Makram OM, Elhady M, et al. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. Epidemiol Infect. 2018;146(13):1625-7.

17. Yang S, Kou SC, Lu F, Brownstein JS, Brooke N, Santillana M. Advances in using Internet searches to track dengue. PLoS Comput Biol. 2017;13(7):e1005607.

18. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Glob Health Action. 2019;12(1):1552652.

19. Ho HT, Carvajal TM, Bautista JR, Capistrano JD, Viacrusis KM, Hernandez LF, et al. Using Google Trends to examine the spatio-temporal incidence and behavioral patterns of dengue disease: a case study in Metropolitan Manila, Philippines. Trop Med Infect Dis. 2018;3(4):118.

20. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google search trends predicting disease outbreaks: an analysis from India. Healthc Inform Res. 2018;24(4):300-8.

21. Devore JL. Probability and statistics for engineering and the sciences. 9th ed. Boston: Cengage Learning; c2016. p. 193-212.

22. R Studio.com. Boston: Learn how BioBuilder teaches aspiring Bioengineers remotely, using RStudio Cloud; 2018 [cited 2019 Feb 15]. Available from: https://www.rstudio.com/

23. Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, et al. Harnessing big data for communicable tropical and sub-tropical disorders: implications from a systematic review of the literature. Front Public Health. 2018;6:90.

24. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res. 2018;20(11):e270.