

O uso do Twitter como minerador de eventos adversos de medicamentos de combate à malária: o caso da doxiciclina

Mining in Twitter for adverse events from malaria drugs: the case of doxycycline

El uso de Twitter como localizador de eventos adversos con medicamentos de combate a la malaria: el caso de la doxiciclina

Felipe Vieira Duval ¹

Fabricao Alves Barbosa da Silva ¹

doi: 10.1590/0102-311X00033417

Resumo

Durante o período de pós-comercialização, quando medicamentos são usados por grandes populações e por períodos de tempo maiores, eventos adversos (EA) inesperados podem ocorrer, o que pode alterar a relação risco-benefício dos medicamentos o suficiente para exigir uma ação regulatória. Eventos adversos são agravos à saúde que podem surgir durante o tratamento com um produto farmacêutico, os quais, no período de pós-comercialização do medicamento, podem requerer um aumento significativo de cuidados de saúde e resultar em danos desnecessários aos pacientes, muitas vezes fatais. Portanto, o quanto antes, a descoberta de EA no período de pós-comercialização é um objetivo principal do sistema de saúde. Alguns países possuem sistemas de vigilância farmacológica responsáveis pela coleta de relatórios voluntários de EA na pós-comercialização, mas estudos já demonstraram que, com a utilização de redes sociais, pode-se conseguir um número maior e mais rápido de relatórios. O objetivo principal deste projeto é construir um sistema totalmente automatizado que utilize o Twitter como fonte para encontrar EA novos e já conhecidos e fazer a análise estatística dos dados obtidos. Para isso, foi construído um sistema que coleta, processa, analisa e avalia tweets em busca de EA, comparando-os com dados da Agência Americana de Controle de Alimentos e Medicamentos (FDA) e do padrão de referência construído. Nos resultados obtidos, conseguimos encontrar EA novos e já existentes relacionados ao medicamento doxiciclina, o que demonstra que o Twitter, quando utilizado em conjunto com outras fontes de dados, pode ser útil para a farmacovigilância.

Controle de Medicamentos e Entorpecentes; Ontologia Biológica; Processamento de Linguagem Natural; Mídias Sociais; Base de Dados

Correspondência

F. V. Duval

Rua 35, Qd 73, Condomínio Colinazul nº 7, Niterói, RJ
24342-086, Brasil.

felipeduval@gmail.com

¹ Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.



Introdução

Durante o período de pós-comercialização, quando medicamentos são usados por grandes populações e por períodos maiores, eventos adversos (EA) podem ocorrer, o que pode alterar a relação risco-benefício dos medicamentos o suficiente para exigir uma ação regulatória. Os EA são definidos como agravos à saúde de um usuário ou de um paciente que podem surgir durante o tratamento com um produto farmacêutico, podendo ser erros de medicação, desvio de qualidade dos medicamentos, reações adversas a medicamentos (RAM), interações medicamentosas e intoxicações ¹.

Segundo a Organização Mundial da Saúde (OMS), a farmacovigilância é definida como “a ciência e atividades relativas à identificação, avaliação, compreensão e prevenção de efeitos adversos ou quaisquer problemas relacionados ao uso de medicamentos” ². Cabe à farmacovigilância identificar, avaliar e monitorar a ocorrência dos EA relacionados ao uso dos medicamentos, com o objetivo de garantir que os benefícios sejam maiores que os riscos por eles causados ¹. Para conseguir esse objetivo, o principal instrumento da farmacovigilância é a notificação espontânea, documento que informa aos órgãos do governo sobre os EA que ocorreram por uso dos medicamentos.

No Brasil, as ações de farmacovigilância são realizadas de forma compartilhada pelas vigilâncias sanitárias dos estados, municípios e pela Agência Nacional de Vigilância Sanitária (Anvisa) ^{2,3}. A taxa de notificações de EA recebidas pela Anvisa é baixa ⁴, sendo muito inferior à meta proposta pela literatura internacional, que sugere trezentas notificações para cada um milhão de habitantes ⁵. Por isso, a utilização de outras fontes de dados para a detecção de EA se faz necessária.

Os EA podem ser identificados durante a fase de estudo sobre o medicamento que ocorre antes da comercialização, conhecida como fase clínica. Os testes clínicos ocorrem em três etapas distintas, conhecidas como fases I, II e III, sendo iniciados com voluntários saudáveis e número limitado de pacientes. Além disso, a seleção e o tratamento dos pacientes geralmente diferem dos utilizados na prática clínica ^{6,7}. EA detectados tardiamente no período de pós-comercialização (também conhecido como fase IV) podem requerer um aumento significativo de cuidados de saúde e resultar em danos desnecessários, muitas vezes fatais, aos pacientes ⁸. Portanto, a descoberta, o quanto antes, de EA no período de pós-comercialização é um objetivo principal do sistema de saúde e, em particular, dos sistemas de vigilância farmacológica.

Métodos computacionais comumente referidos como “detecção de sinais” permitem que os avaliadores de segurança de medicamentos analisem grandes volumes de dados para identificar sinais de risco de potenciais EA, e também provaram ser um componente fundamental na farmacovigilância. Como exemplo, a Agência Americana de Controle de Alimentos e Medicamentos (FDA), usa rotineiramente um processo de rastreamento de sinais para calcular estatísticas, relatando associações para todos os milhões de combinações de medicamentos e eventos em seu sistema de comunicações de EA ⁸. Esses sinais, por si sós, não são suficientes para estabelecer uma relação causal, mas são considerados avisos iniciais que requerem avaliação aprofundada por especialistas para estabelecer a causalidade.

Dedicar recursos para a posterior avaliação de cada um dos múltiplos sinais normalmente gerados por algoritmos de detecção não é viável. Recursos desviados para indicações falsas podem inviabilizar um sistema de vigilância farmacológica ⁹. Portanto, estratégias automatizadas para reduzir a quantidade de falsos positivos e definir prioridades, de modo a permitir que apenas os sinais mais promissores sejam avaliados, são imperativas.

Desta forma, a contribuição principal deste artigo é a proposta do TweetAEMiner (Tweet Adverse Event Miner), um sistema automatizado de vigilância farmacológica capaz de identificar associações novas e já existentes de “medicamento-EA”, com o uso de mineração de texto.

A mineração de textos consiste em usar técnicas para recuperar informação textual, extrair informação, bem como processar linguagem natural com algoritmos e métodos de descoberta de conhecimento, mineração de dados e aprendizado de máquina ¹⁰.

Neste trabalho, o Twitter foi utilizado como fonte para a mineração de texto. Trata-se de uma base de dados não convencional, devido a maior facilidade e rapidez de acesso aos seus dados. Exemplos de outras bases não convencionais que vêm sendo utilizadas recentemente em vigilância epidemiológica são logs de busca ^{11,12,13} e redes sociais ^{14,15}.

A maioria dos trabalhos anteriores de mineração de texto relacionados com a farmacovigilância está focada em registros eletrônicos de saúde e em relatos de casos médicos ^{16,17}. Harpaz et al. ¹⁸

fornece um estudo aprofundado sobre as abordagens existentes para a fase de pós-comercialização, explorando vários recursos, tais como registros eletrônicos de saúde e sistemas de relato espontâneo de EA. As redes sociais vêm sendo utilizadas para esse propósito recentemente. Leaman et al.¹⁹ analisaram os comentários de usuários em redes sociais e demonstraram que há neles informações sobre medicamentos que podem ser extraídas para posterior análise. Em um estudo recente, Yates & Goharian²⁰ analisaram o valor dos comentários de usuários em revelar EA desconhecidos, avaliando as RAM extraídas com a base de dados SIDER (<http://sideeffects.embl.de/>), que contém informações sobre os EA conhecidos²¹.

A grande maioria das pesquisas que utilizam o Twitter como fonte de dados e que são voltadas para a área médica busca outras informações que não EA. Alguns estudos utilizaram o Twitter para essa função^{22,23,24} e mostraram que o uso de seus *tweets* pode levar à farmacovigilância em tempo real. Freifeld et al.²³ utilizou o Twitter para avaliar o nível de concordância entre os *tweets* com menções de EA (Proto-AE – *posts with resemblance to adverse events*) e relatos espontâneos de EA da FDA (FAERS – FDA Adverse Event Reporting System). Nesse estudo, foram coletados 6,9 milhões de *tweets* com nomes de medicamentos, dos quais 4.401 foram identificados como Proto-AE. Eles mostraram que o Twitter teve quase três vezes mais Proto-AE do que relatos da FDA²³.

As pesquisas que buscam EA no Twitter geralmente coletam dados de alguns poucos meses para encontrar RAM conhecidas, usam uma ou nenhuma ontologia (modelo de dados que representa um conjunto de conceitos e relacionamentos dentro de um domínio) para fazer isso e têm etapas manuais do seu *pipeline* (sequência de operações em que a saída de uma etapa/operação serve de entrada para a próxima operação na sequência). Nesse artigo, é utilizado um *pipeline* automático para coleta, armazenamento e processamento de *tweets* que utiliza uma ontologia completa e totalmente voltada para a busca de RAM.

Devido a limitações do número de palavras que podem ser buscadas no Twitter, este estudo foi focado em RAM, no combate à malária, que foi a doença negligenciada com mais *tweets* no ano de 2014. Dentre esses medicamentos, foi feita uma análise dos EA da doxiciclina encontrados nos *tweets* com valores consolidados de notificação de EA recebidos pela FDA. Entretanto, cabe ressaltar que o sistema descrito neste artigo pode ser adaptado para monitorar múltiplas doenças e medicamentos simultaneamente.

Materiais e métodos

O TweetAEMiner realiza continuamente a coleta de *tweets* com uso da API (*application programming interface*) do Twitter, com palavras pré-determinadas (doenças ou medicamentos). Esses *tweets* são armazenados na base de dados. Periodicamente o sistema inicia o processamento e a análise desses *tweets*. Atualmente o sistema está configurado para rodar o processamento e análise aos domingos, pois é quando começa uma nova semana no calendário epidemiológico²⁵, mas essa periodicidade pode ser facilmente alterada, caso seja necessário. No processamento dos *tweets*, é utilizado um processador de linguagem natural (PLN) e, com a saída desse processamento, é feita a análise estatística dos dados. Por último, é feita uma avaliação dos resultados obtidos com um padrão de referência.

O sistema gera uma lista de sinais específicos, que são avaliados com base em um padrão de referência. Um sinal corresponde a uma associação “medicamento-EA” identificada pelo *pipeline*.

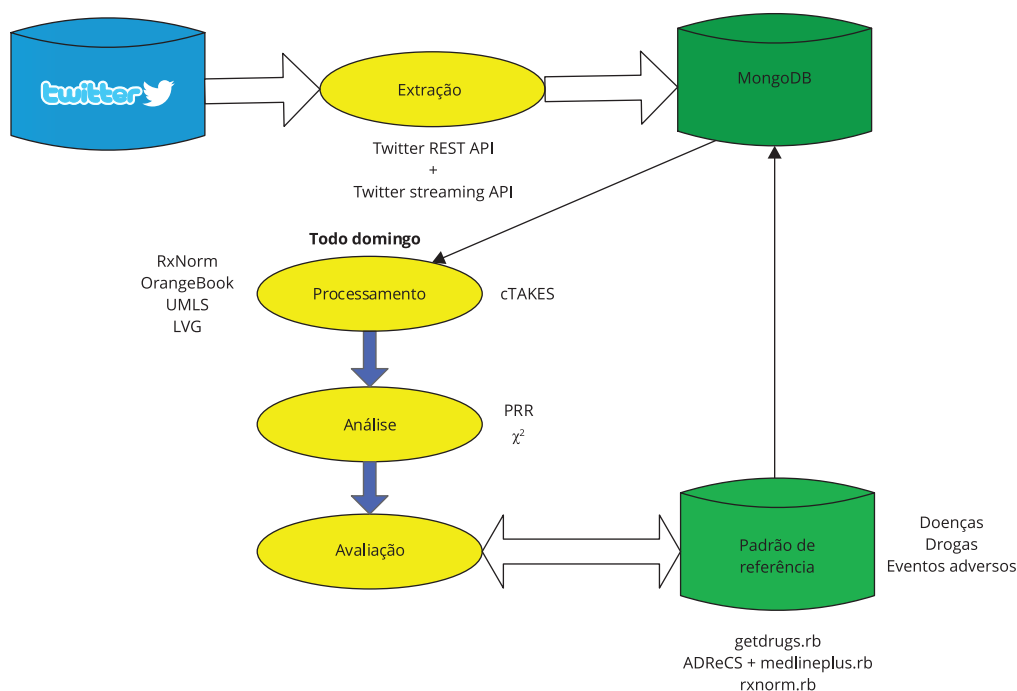
Na Figura 1, são mostradas as quatro etapas do *pipeline*: extração, processamento, análise e avaliação dos dados. Além das etapas, a Figura 1 também mostra a base de dados utilizada para armazenar os *tweets* e o padrão de referência criado.

Extração

O Twitter possui duas API para coleta de *tweets*: a REST API (<http://dev.twitter.com/rest/public>) e a Streaming API (<http://dev.twitter.com/docs/api/streaming>). As duas API permitem apenas o acesso a *tweets* recentes, portanto os efetivamente coletados serão úteis para pesquisas futuras. A coleta do material está sendo realizada desde o início de 2014, com a utilização das API citadas.

Figura 1

Metodologia do TweetAEMiner.



Nota: *pipeline* do sistema. Na cor amarela, as quatro etapas do processo; na cor verde, as bases utilizadas para armazenamento dos *tweets* e como padrão de referência; na cor azul, o Twitter.

Como uma abordagem inicial, foram coletados *tweets* relacionados a doenças negligenciadas, como malária, dengue, doença de chagas, tuberculose e leishmaniose ²⁶. Posteriormente, as consultas foram expandidas para outras doenças, incluindo também as não negligenciadas, como a aids.

Uma análise preliminar dos dados coletados indicou que malária era a doença com mais *tweets*. Apesar de algumas dessas doenças não terem ainda um medicamento associado, as mensagens referentes a elas podem ser úteis em outros projetos, como, por exemplo, projetos de estudos epidemiológicos.

Dada a limitação do número de palavras que podem ser buscadas na rede social em questão, coletaram-se apenas *tweets* de medicamentos utilizados no combate à malária.

Para obter todos os nomes de medicamentos relacionados com a malária, foi utilizado o *site* <http://www.drugs.com>. No *site*, é possível encontrar nomes de medicamentos comerciais e genéricos. Para facilitar o levantamento desses dados, foi desenvolvido um programa que relaciona os medicamentos associados ao nome de cada doença. Para a doença malária, foram utilizados 19 medicamentos, sendo dez nomes comerciais (Plaquenil, Malarone, Doryx, Lariam, Daraprim, Aralen, Fansidar, Morgidox, Ocudox e Oraxyl) e nove nomes genéricos (atovaquone, proguanil, doxycycline, mefloquine, pyrimethamine, sulfadoxine, hydroxychloroquine, chloroquine e primaquine). Dentre esses medicamentos, o que apresentou a maior quantidade de *tweets* em 2014 foi a doxiciclina, como pode ser visto na Tabela 1, sendo por isso escolhida como alvo da análise.

O TweetAEMiner foi feito de maneira a possibilitar a portabilidade do pipeline para outros tipos de textos além de *tweets*, com o mínimo de esforço possível. Para isso, bastaria adequar a parte da extração para alguma outra fonte de textos que não fosse o Twitter.

Tabela 1Números de *tweets* citando medicamentos no combate à malária em 2014.

Medicamentos	n
Morgidox	0
Ocudox	0
Oraxyl	0
Daraprim	35
Sulfadoxine	61
Proguanil	98
Aralen	122
Doryx	173
Atovaquone	191
Fansidar	193
Pyrimethamine	216
Primaquine	359
Lariam	671
Hydroxychloroquine	819
Malarone	890
Plaquenil	982
Mefloquine	1.312
Chloroquine	2.912
Doxycycline	14.333

Padrão de referência

O padrão de referência foi construído de modo a ser uma base de dados amplamente aceita com todos os EA atualmente conhecidos. Para tanto, foram utilizados principalmente o *Adverse Drug Reaction Classification System* (ADReCS) ²⁷, uma ontologia para termos de reações adversas que usa fontes médicas. Foi acrescentado o relacionamento entre doenças e seus medicamentos a essa ontologia.

Com essas fontes, foi criada uma base com as doenças estudadas, os medicamentos utilizados em seus tratamentos e os EA de cada uma delas.

Atualmente, estão sendo processados apenas os *tweets* em inglês, pois todas as fontes utilizadas no padrão de referência são constituídas somente de palavras da língua inglesa.

Processamento

Depois do processo de extração, os *tweets* passam por um PLN. Existem algumas ferramentas PLN utilizadas no campo da medicina, como Medlee ²⁸, cTAKES ²⁹ e MetaMap ³⁰. O cTAKES foi escolhido por ser um PLN de código aberto usado para extrair informações a partir de texto livre. Ele utiliza diferentes vocabulários de diversas fontes médicas.

O cTAKES é utilizado em um programa que processa os *tweets* armazenados, gerando como saída as doenças, os medicamentos e as reações adversas associadas, bem como outras informações médicas que foram encontradas no texto.

Apesar de o TweetAEMiner utilizar *tweets* em vez de relatos espontâneos, filtram-se as mensagens para que se tenha, no mínimo, um medicamento e um EA, descartando-se aquelas que não os tenham. Trata-se de uma abordagem parecida com os Proto-AE de Freifeld et al. ²³.

Neste trabalho, é utilizada uma abordagem que se baseia na seleção de medicamentos (*drug-based approach*) ³¹. Essa abordagem foi escolhida por não se saber o número de *tweets* com determinado EA, bem como para determinar a quantidade de *tweets* com EA e os medicamentos relacionados à

doença-alvo. Com essa abordagem, é mais apropriado considerar um *tweet* com o nome do medicamento do que coletar um qualquer que pode não estar relacionado a medicamentos.

Análise

Após o processamento dos *tweets*, é utilizada uma medida de análise de desproporcionalidade para os dados a serem analisados. Métodos de análise de desproporcionalidade (ADP) em vigilância farmacológica representam a classe principal de métodos analíticos para sistemas de relato espontâneo (SRE) ¹⁸. SRE são relatos que compreendem um ou mais medicamentos, de um ou mais EA e, possivelmente, alguns dados demográficos de base. Esses métodos identificam associações relevantes em bases de dados de SRE, com foco em projeções de baixa dimensionalidade dos dados, mais especificamente tabelas de contingência 2x2. Tanto a FDA como a OMS utilizam métodos de ADP para encontrar essas associações ¹⁸. Essa medida foi usada para classificar pares “medicamento-EA” identificados na etapa de processamento anterior. Deve-se observar que o método de análise pode variar em função dos dados que forem processados. SRE baseados em RAM realizam a detecção de sinais mais frequentemente utilizando medidas de desproporcionalidade.

A tarefa básica de um método ADP é a classificação das tabelas em ordem de “interesse”. Diferentes métodos ADP focam em diferentes medidas estatísticas de associação como a sua medida de “interesse”. A Tabela 2 apresenta as fórmulas para as diferentes medidas de associação mais comumente usadas, juntamente com a sua interpretação probabilística, em que “-medicamento” denota os relatos que não incluem o medicamento-alvo.

Um medicamento em particular que causa um EA específico mais que qualquer outro normalmente terá o valor da medida de associação mais elevado. Se um EA e um medicamento são estocasticamente independentes, o valor da medida de associação receberá o valor igual a 1. Como normalmente cada EA de um medicamento individual ocorre em uma proporção pequena do total de notificações, geralmente temos $a \ll b$ ou $a \ll c$ e $c \ll d$ e, na prática, essas medidas tendem a ter valores e interpretações idênticos. Por exemplo, um valor de 3 indica que há três vezes mais notificações envolvendo o par medicamento-EA do que esperado, se não houvesse associação entre os dois ³².

Tabela 2

Medidas comuns de associação em análises de sistemas de relato espontâneo (SRE).

Medida de associação	Fórmula	Valor	Interpretação probabilística
Relative reporting ratio (RRR)	$\frac{t \cdot a}{m \cdot n}$	35,57355	$\frac{Pr(EA medicamento)}{Pr(EA)}$
Proportional reporting ratio (PRR)	$\frac{(a(t-n))}{c \cdot n}$	37,36421	$\frac{Pr(EA medicamento)}{Pr(EA medicamento)}$
Reporting odds ratio (ROR)	$\frac{a \cdot d}{c \cdot b}$	37,57431	$\frac{Pr(EA medicamento)}{Pr(EA medicamento)}$ $\frac{Pr(EA medicamento)}{Pr(EA medicamento)}$
Information component	$\log_2(RRR)$	5,12573	$\log_2 \frac{Pr(EA medicamento)}{Pr(EA)}$

EA: eventos adversos.

Nota: as letras “a”, “b”, “c”, “d” são os valores da tabela de contingência 2x2 para um medicamento e um EA. As letras “m”, “n” e “t” são somas, conforme exemplificado em Duval et al. ²⁶.

Para calcular as associações, é utilizada a abordagem frequentista *proportional reporting ratio* (PRR) para a análise de desproporcionalidade. Medidas bayesianas tendem a produzir valores menos extremos do que PRR, quando o número de casos é muito pequeno. No entanto, na ocasião em que a sensibilidade, a especificidade e o poder preditivo dessas medidas foram comparados com dados holandeses em 2002³³, não foram encontradas diferenças importantes quando, pelo menos, três casos foram relatados. Além disso, cabe ressaltar que a medida PRR já foi utilizada em vários trabalhos para detecção de RAM em SER^{32,34,35} e é uma das principais medidas utilizadas pela União Europeia. Juntamente com o PRR, foi calculado o seu intervalo de 95% de confiança (IC95%) e também realizado o teste χ^2 para validação dos sinais gerados, da mesma maneira como é realizada pelo SRE utilizado pela União Europeia, o EudraVigilance³⁴.

Avaliação

O TweetAEMiner verifica se, na análise dos dados, houve algum sinal (uma associação “medicamento-EA”) de maneira similar ao EudraVigilance, com a medida de desproporcionalidade calculada, PRR, em conjunto com seu IC95% e também com a utilização do teste χ^2 .

O PRR, por se tratar de um método muito sensível, pode gerar muitos falsos positivos, especialmente se o número de notificações for baixo. Para reduzir isso, um dos critérios utilizados é o cálculo do IC95%.

O IC95% para o logaritmo neperiano de PRR é estimado como $\bar{p} \pm se$, em que “se” (*standard error*) é o erro padrão da média do logaritmo natural do PRR^{33,36}. Se o PRR for apresentado com intervalo de confiança de 95%, será considerado como um sinal de desproporcionalidade quando³⁴: Limite inferior do intervalo ≥ 1 ; número de casos ≥ 3 .

Outra medida para detecção de sinais utilizada em conjunto com o PRR é a estatística do χ^2 , que é um teste de independência de variáveis categóricas utilizado como uma medida alternativa da heterogeneidade da tabela de contingência construída com um medicamento M e um EA³⁴.

Se o PRR for apresentado com a medida χ^2 , será considerado como um sinal de desproporcionalidade quando: PRR ≥ 2 ; $\chi^2 \geq 4$; número de casos ≥ 3 .

Além da análise dos *tweets*, também foi feita a análise de dados da FDA para uma comparação dos sinais gerados em ambas. Os sinais detectados em cada uma das análises foram agrupados em três tipos:

- (a) Tipo A: gerados pelo critério do intervalo de confiança do PRR, ou seja, quando o limite inferior do IC95% do PRR for ≥ 1 e a quantidade de *tweets*/relatos for ≥ 3 ;
- (b) Tipo B: gerados pelo critério do χ^2 , ou seja, PRR ≥ 2 e $\chi^2 \geq 4$ e a quantidade de *tweets*/relatos for ≥ 3 ;
- (c) Tipo C: quando ocorrerem os sinais do tipo A e B.

Resultados

Um dos principais resultados do artigo foi a construção de uma ferramenta automática para a coleta e análise de EA no Twitter. Dentre os 19 medicamentos no combate à malária que foram utilizados como filtro dos *tweets*, a doxiciclina foi a que possuiu a maior quantidade de mensagens, como mostrado na Tabela 1, sendo, por isso, a escolhida para análise. Na avaliação dos resultados, foi feita uma comparação da análise dos dados obtidos pelo TweetAEMiner com dados da FDA obtidos por meio do site <https://open.fda.gov/>.

Análise de dados do Twitter

No cálculo da análise de desproporcionalidade, foi utilizada a medida PRR, considerando somente os *tweets* que tivessem algum EA. Foram utilizados também todos os sinônimos de RAM existentes no ADReCS na contagem para a construção das tabelas de contingência.

Na Tabela 3, é mostrado o relatório do PRR para o medicamento doxiciclina com os EA conhecidos do medicamento no padrão de referência e que possuíam, ao menos, um *tweet*.

Tabela 3Relatório do *proportional reporting ratio* (PRR) para eventos adversos (EA) do medicamento doxiciclina (Twitter).

EA	PRR(-) *	PRR **	PRR(+) ***	χ^2	Tweets	FDA #
Desconforto abdominal	Não calculado	99,9	Não calculado	2,356	11	
Distensão abdominal	Não calculado	99,9	Não calculado	1,071	5	
Dor abdominal supertior	Não calculado	99,9	Não calculado	6,434	30	
Abscesso	Não calculado	99,9	Não calculado	0,428	2	
Anemia	0,197	1,634	13,568	0,166	6	
Reação anafilática	0,038	0,272	1,933	1,529	2	SIM
Angioedema	0,108	0,233	0,504	12,807	12	
Anorexia	0,017	0,272	4,353	0,764	1	
Ansiedade	1,812	4,466	11,007	10,022	82	SIM
Estomatite aftosa	1,079	4,493	18,716	4,035	33	
Artralgia	0,314	0,953	2,894	0,006	14	
Dor lombar	0,427	0,657	1,012	2,897	70	
Pressão arterial aumentada	Não calculado	99,9	Não calculado	2,356	11	
Bronquite	0,113	0,272	0,654	7,651	10	
Candidíase	Não calculado	99,9	Não calculado	2,999	14	
Tosse	0,567	1,634	4,706	0,664	24	
Apetite diminuído	Não calculado	99,9	Não calculado	0,428	2	
Dermatite	Não calculado	99,9	Não calculado	0,214	1	SIM
Diarreia	0,214	0,681	2,169	0,336	10	
Desconforto	0,017	0,272	4,353	0,764	1	SIM
Dispepsia	Não calculado	99,9	Não calculado	0,642	3	
Disfagia	0,055	0,272	1,349	2,293	3	
Infecção no ouvido	0,113	0,272	0,654	7,651	10	
Angústia emocional	Não calculado	99,9	Não calculado	0,214	1	SIM
Infecção por fungos	1,417	4,539	14,543	6,159	50	
Gengivite	Não calculado	99,9	Não calculado	0,214	1	
Anemia hemolítica	0,017	0,272	4,353	0,764	1	
Dor de cabeça	0,165	0,327	0,648	8,937	18	
Hipersensibilidade	0,482	0,754	1,179	1,211	72	
Hipertensão	Não calculado	99,9	Não calculado	1,713	8	
Infecção	2,664	4,341	7,076	32,958	271	
Inflamação	Não calculado	99,9	Não calculado	1,499	7	
Influenza	0,229	0,256	0,285	528,852	557	
Lesão	0,172	0,363	0,767	6,032	16	SIM
Insônia	0,088	0,182	0,377	20,92	12	
Pressão intracraniana aumentada	Não calculado	99,9	Não calculado	0,214	1	
Leucopenia	Não calculado	99,9	Não calculado	0,428	2	
Mal-estar	Não calculado	99,9	Não calculado	0,856	4	SIM
Espasmos musculares	Não calculado	99,9	Não calculado	2,356	11	SIM
Mialgia	0,085	0,817	7,852	0,024	3	
Congestão nasal	Não calculado	99,9	Não calculado	0,214	1	
Nasofaringite	0,009	0,091	0,872	5,37	1	
Nausea	0,943	3,949	16,54	3,245	29	
Edema	0,009	0,091	0,872	5,37	1	
Úlcera do esôfago	Não calculado	99,9	Não calculado	0,642	3	SIM
Esofagite	Não calculado	99,9	Não calculado	0,642	3	
Dor orofaríngea	0,039	0,163	0,683	6,316	3	
Dor	1,556	2,465	3,905	12,485	181	

(continua)

Tabela 3 (continuação)

EA	PRR(-) *	PRR **	PRR(+) ***	χ^2	Tweets	FDA #
Reação fotossensível	Não calculado	99,9	Não calculado	1,928	9	SIM
Transtorno da pigmentação	0,049	0,545	6,005	0,199	2	
Erupção cutânea	0,974	2,451	6,17	3,048	45	
Rinorréia	Não calculado	99,9	Não calculado	0,214	1	
Sinusite	0,172	0,272	0,432	27,638	36	
Síndrome de Stevens-Johnson	0,036	0,091	0,229	32,26	6	
Estomatite	Não calculado	99,9	Não calculado	0,428	2	
Inchaço	1,383	10,076	73,414	6,272	37	
Tensão	Não calculado	99,9	Não calculado	4,716	22	
Trombocitopenia	Não calculado	99,9	Não calculado	0,428	2	
Abscesso dentário	0,038	0,272	1,933	1,529	2	
Dor de dente	Não calculado	99,9	Não calculado	1,499	7	
Úlcera	Não calculado	99,9	Não calculado	3,643	17	
Urticária	0,064	0,117	0,213	55,349	15	
Vômito	Não calculado	99,9	Não calculado	3,428	16	SIM

FDA: Agência Americana de Controle de Alimentos e Medicamentos.

Nota: quando é detectado um sinal pelo χ^2 , a célula é preenchida na cor vermelha; quando detectado um sinal pelo intervalo de 95% confiança (IC95%) do PRR, a célula é preenchida na cor laranja. A coluna "FDA" é preenchida na cor verde quando o sinal tiver ocorrido no Twitter e na FDA.

* Limite inferior do IC95% do PRR;

** Valor do PRR para o EA;

*** Limite superior do IC95% do PRR;

Mostra se houve sinal desse EA na FDA no mesmo período de 2014.

Em algumas situações, quando o número de tweets com o medicamento-alvo e o EA em análise é maior que zero e o número de tweets com o EA mas sem o medicamento-alvo é igual a zero, o PRR não pode ser calculado. Por esse motivo, é atribuído arbitrariamente "99,9" na coluna "PRR" da Tabela 3 para refletir a presença de um possível sinal. Nesses casos, os limites do intervalo de confiança não são calculados, como pode ser visto nas colunas "PRR(-)" e "PRR(+)".

Foram detectados sinais para dois possíveis novos EA: alopecia e rosácea. Ambos também aparecem no FDA no mesmo período, como pode ser visto na Tabela 4. No FDA, são relatados mais de 200 EA.

Análise dos dados da FDA

A análise dos dados da FDA é feita de maneira similar à do Twitter, porém com uso dos relatos da FDA, no mesmo período, com os 19 medicamentos.

Diferentemente do que ocorre na rede social consultada, o medicamento com maior quantidade de relatos no FDA foi a hidroxicloriguina. Apenas o medicamento Oraxyl não teve nenhum relato no ano de 2014 (Tabela 5). Como os relatos são voltados especificamente para a detecção de EA, é normal que sua análise produza um grande número de sinais. A doxiciclina, por exemplo, teve relatos com mais de 200 EA diferentes, e desses, 138 geraram sinais.

Geração de sinais tipo A, B e C

Não houve sinais do tipo A gerados pelo Twitter. A FDA gerou um total de 51 sinais do tipo A, dos quais 40 não se encontram no padrão de referência. Os 11 EA dos sinais que estavam no padrão de referência são dor abdominal, desconforto, hipersensibilidade, mal-estar, espasmos musculares, mialgia, náusea, erupção cutânea, exantema eritematoso, urticaria e vômito.

Foram gerados dois sinais do tipo B pelo Twitter, para os EA dor abdominal superior e tensão, ambos presentes no padrão de referência. Também foram gerados outros dois sinais do tipo B que

Tabela 4

Comparativo entre quantidade de eventos adversos (EA) encontrados nos *tweets* e nos relatórios da Agência Americana de Controle de Alimentos e Medicamentos (FDA) para o medicamento doxiciclina, no ano de 2014.

EA	<i>Tweets</i>	Relatórios FDA
Desconforto abdominal	11	21
Distensão abdominal	5	10
Dor abdominal supertior	30	32
Abscesso	2	-
Alopecia	155	18
Anemia	6	33
Reação anafilática	2	12
Angioedema	12	-
Anorexia	1	-
Ansiedade	82	86
Estomatite aftosa	33	-
Artralgia	14	48
Dor lombar	70	29
Pressão arterial aumentada	11	16
Bronquite	10	33
Candidiase	14	-
Tosse	24	48
Apetite diminuído	2	36
Dermatite	1	11
Diarreia	10	96
Desconforto	1	17
Dispepsia	3	11
Disfagia	3	17
Infecção no ouvido	10	-
Angústia emocional	1	47
Infecção por fungos	50	-
Gengivite	1	-
Anemia hemolítica	1	-
Dor de cabeça	18	119
Hipersensibilidade	72	29
Hipertensão	8	22
Infecção	271	19
Inflamação	7	14
Influenza	557	16
Lesão	16	54
Insônia	12	29
Pressão intracraniana aumentada	1	-
Leucopenia	2	-
Mal-estar	4	91
Espasmos musculares	11	42
Mialgia	3	32
Congestão nasal	1	-
Nasofaringite	1	20
Nausea	29	200
Edema	1	12
Úlcera do esôfago	3	18

(continua)

Tabela 4 (continuação)

EA	Tweets	Relatórios FDA
Esofagite	3	-
Dor orofaríngea	3	23
Dor	181	122
Reação fotosssensível	9	18
Transtorno da pigmentação	2	-
Erupção cutânea	45	90
Rinorréia	1	15
Rosácea	27	9
Sinusite	36	18
Síndrome de Stevens-Johnson	6	-
Estomatite	2	-
Inchaço	37	9
Tensão	22	-
Trombocitopenia	2	12
Abscesso dentário	2	-
Dor de dente	7	-
Úlcera	17	-
Urticária	15	47
Vômito	16	137

Tabela 5

Quantidade de eventos adversos (EA) nos relatos com medicamentos no combate à malária em 2014.

Medicamentos	n
Oraxyl	0
Primaquine	24
Fansidar	34
Sulfadoxine	36
Aralen	48
Lariam	110
Daraprim	128
Pyrimethamine	198
Mefloquine	319
Malarone	385
Proguanil	429
Morgidox	533
Ocudox	533
Chloroquine	621
Doryx	640
Atovaquone	1.040
Doxycycline	6.079
Plaquenil	7.664
Hydroxychloroquine	10.564

não estão no padrão de referência para os EA alopecia e rosácea. Dentre esses sinais, apenas rosácea também ocorreu na FDA, que teve um total de 24 sinais do tipo B, dos quais apenas menorragia se encontra no padrão de referência.

O Twitter gerou um total de seis sinais do tipo C para os EA: ansiedade, estomatite aftosa, infecção fúngica, infecção, dor e inchaço. Todos estão presentes no padrão de referência de EA para doxiciclina. Desses sinais, apenas ansiedade ocorreu na FDA, que teve um total 63 sinais, dos quais oito estavam presentes no padrão de referência: reação anafilática, ansiedade, dermatite, angústia emocional, lesão, úlcera de esôfago, reação de fotossensibilidade e erupção maculopular e outros 55 sinais que não se encontravam no padrão de referência.

Discussão

Com o intuito de construir um sistema capaz de coletar, armazenar e processar tweets relacionados a medicamentos, foi implementado inicialmente um coletor utilizando as API do próprio Twitter. Como essa API não possibilita a aquisição de mensagens antigas, o TweetAEMiner já vem coletando *tweets* com diversos medicamentos e doenças que não foram alvo deste artigo para utilização em trabalhos futuros.

A doença com o maior número de *tweets* foi a dengue, mas, por esta não possuir medicamentos para o seu tratamento, focou-se o estudo-teste da ferramenta na doença malária e em seus medicamentos, pois era a segunda enfermidade com maior quantidade de mensagens.

Foram coletados *tweets* durante todo o ano de 2014 com medicamentos relacionados à malária. Dentre esses medicamentos, alguns não apresentaram qualquer *tweet*, como Morgidox, Ocudox e Oraxyl. O maior número de *tweets* retornados foi do medicamento doxiciclina (14.333 *tweets* sem incluir medicamentos similares), como mostrado na Tabela 4. Outros medicamentos não apresentaram uma quantidade significativa de mensagens para que fosse feita alguma análise, ou então não possuíam qualquer EA associado.

Não há consenso sobre qual é a melhor abordagem para a análise de desproporcionalidade: frequentista ou bayesiana³⁷. Do ponto de vista do cenário internacional, ambas as abordagens são utilizadas. A FDA utiliza o *Multi-Item Gamma-Poisson Shrinker* (MGPS)¹⁸, um método bayesiano. Na União Europeia, era utilizado o método frequentista PRR na época em que foi feita a análise, sendo atualmente utilizado o método *reporting odd ratios* (ROR). A OMS, por sua vez, utiliza o *Bayesian Confidence Propagation Neural Network* (BCPNN)¹⁸, que é uma versão bayesiana do *information component*. Em função dessas observações, optou-se por realizar a primeira análise com o PRR, por ser mais simples, se comparado aos outros métodos.

Na análise do Twitter, foram detectados sinais para oito EA conhecidos da doxiciclina: dor abdominal superior, ansiedade, estomatite aftosa, infecção por fungos, infecção, dor, inchaço e tensão. Além desses, foram detectados dois EA ainda não relacionados à doxiciclina: alopecia e rosácea. Dos oito EA conhecidos da doxiciclina detectados por meio da análise dos *tweets*, apenas ansiedade também foi verificado na análise de dados da FDA. Seria interessante realizar essa comparação por um período maior de tempo para verificar se os sinais gerados pelo Twitter para esses oito EA tendem a aumentar, se manter ou diminuir. Caso esses sinais continuem a aparecer somente nos *tweets*, é um indício potencial de que as pessoas utilizam mais essa rede social do que relatórios formais para reportar EA.

Comparando-se as Tabelas 3 e 4, verificou-se a existência de três EA presentes no padrão de referência e que geraram sinais somente no Twitter, pois não houve relatos da FDA associados. São eles: estomatite aftosa, infecção fúngica e tensão. Isso demonstra que EA que não aparecem nos relatos poderiam ser detectados no Twitter, já que também são EA de doxiciclina.

Pesquisando sobre os dois EA que não estavam no padrão de referência (rosácea e alopecia) e que foram detectados pelo Twitter, verificou-se que eles também aparecem nos relatos do FDA do mesmo período. Existem tanto relatos de que a doxiciclina poderia causar a calvície como também que poderia evitá-la. Sobre o EA rosácea, a grande maioria de *tweets* e de relatos pesquisados indicava que o medicamento era utilizado para o seu tratamento, e não que era o responsável pela sua causa³⁸.

Tanto alopecia como rosácea aparecem nos relatos da FDA, contudo apenas rosácea gerou um sinal na análise de dados. Isso é mais um indício de que a utilização de múltiplas fontes de dados traz uma maior sensibilidade ao sistema de detecção automática de sinais, pois, se forem considerados apenas eventos raros, a análise de dados de múltiplas fontes é necessária para se conseguir o poder estatístico e a heterogeneidade populacional necessários para detectar diferenças da efetividade de drogas em subpopulações, levando-se em conta diferenças genéticas, étnicas e clínicas ³⁹.

O fato de a alopecia não estar no padrão de referência significa que ela pode ser um potencial novo EA. Além disso, esse sinal não foi detectado pela FDA, mas somente pelo Twitter, indicando que essa rede poderia detectar sinais que escapariam de outras fontes.

Cabe ressaltar que todos os resultados das análises são sinais, e não afirmações sobre relação de causa e efeito entre o medicamento e o EA. Essas afirmações de maneira alguma podem vir a ser feitas de forma automática, sendo necessários estudos posteriores conduzidos por especialistas que viriam a se utilizar de tais sinais como avisos iniciais para justificar uma avaliação mais aprofundada.

É importante enfatizar que os valores de PRR e χ^2 são medidas de associação, e não de causalidade. Por isso, alguns EA podem não ter gerado sinais, apesar de serem relacionados com os medicamentos analisados, e isso ocorre tanto na análise do Twitter, como na da FDA. Nenhuma das duas análises gerou sinais para todos os EA existentes no padrão de referência, como pode ser visto na Tabela 3.

Apesar dos relatos da FDA serem voltados justamente para encontrar RAM, a grande maioria dos 138 sinais gerados era de EA não pertencentes ao padrão de referência (40 do tipo A, 23 do tipo B e 55 do tipo C). Ou seja, apenas 20 EA já se encontravam no padrão de referência relacionados a doxiciclina.

Os resultados obtidos corroboram a ideia de que o Twitter seja útil para a farmacovigilância, mas não como uma fonte de dados isolada, e sim como uma fonte complementar de dados. A rede social foi capaz de gerar tanto sinais novos como os que já existem no padrão de referência, além de sinais que não foram obtidos por meio da análise dos dados da FDA.

Uma crença emergente na pesquisa em farmacovigilância é que a combinação de informações de múltiplas fontes de dados pode levar a descoberta mais eficaz e precisa de EA ⁸. Dependendo das fontes de dados utilizadas e do modo como elas são combinadas, acredita-se que o sistema resultante poderia levar ao aumento da significância estatística dos resultados ou facilitaria novas descobertas que não são possíveis com base em fontes de dados isoladas. Essa hipótese foi preliminarmente confirmada recentemente ⁸, mas novos estudos são necessários. A utilização de múltiplas *pipelines*, com as etapas de processamento, avaliação e análise dos dados, cada um com fontes de dados diferentes, seria uma maneira de corroborar tal hipótese e é uma direção futura de pesquisa importante.

Além de corroborar como mais uma fonte, outro fator importante é a disponibilidade dos dados do Twitter, que permite um acesso em tempo real para a análise dos dados, ao passo que as redes de vigilância farmacológica costumam demorar a disponibilizar os seus dados. A FDA, por exemplo, disponibiliza dados por trimestre, mas não necessariamente esses dados se tornam públicos depois de decorridos três meses. Geralmente as informações relativas aos meses de janeiro, fevereiro e março de um ano só vêm a se tornar públicas depois da metade do trimestre.

Pode-se considerar que a análise dos dados do Twitter se mostrou útil para compor um sistema de farmacovigilância mais completo. Por meio da análise desses dados, foram detectados EA que não se encontravam no padrão de referência (alopecia e rosácea) e, dentre eles, alopecia não se encontrava nos sinais gerados pela FDA. Contudo, mais análises devem ser feitas para corroborar esses resultados, de modo a incluir outros medicamentos e outros períodos de tempo. Também seria interessante realizar uma análise com base em outro método, como o MGPS, que é utilizado pela FDA.

Colaboradores

F. V. Duval participou da análise e interpretação dos dados, redação do artigo e é responsável por todos os aspectos do trabalho na garantia da exatidão e integridade de qualquer parte da obra. F. A. B. Silva colaborou na concepção e projeto do artigo, revisão crítica relevante do conteúdo intelectual e aprovação final da versão a ser publicada.

Informações adicionais

ORCID: Felipe Vieira Duval (0000-0003-4476-1277); Fabrício Alves Barbosa da Silva (0000-0002-8172-5796).

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo suporte financeiro.

Referências

1. Mendes M, Pinheiro R, Avelar K, Teixeira J, Silva G. História da farmacovigilância no Brasil. *Rev Bras Farm* 2008; 89:246-51.
2. World Health Organization. Pharmacovigilance. https://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/ (acessado em 01/Out/2018).
3. Balbino EE, Dias MF. Farmacovigilância: um passo em direção ao uso racional de plantas medicinais e fitoterápicos. *Rev Bras Farmacogn* 2010; 20:992-1000.
4. Mota DM. Evolução e resultados do sistema de farmacovigilância do Brasil [Dissertação de Mestrado]. Porto Alegre: Faculdade de Medicina, Universidade Federal do Rio Grande do Sul; 2017.
5. Meyboom RH, Egberts AC, Gribnau FW, Hekster YA. Pharmacovigilance in perspective. *Drug Saf* 1999; 21:429-47.
6. Venulet J, ten Ham M. Methods for monitoring and documenting adverse drug reactions. *Int J Clin Pharmacol Ther* 1996; 34:112-29.
7. Cardoso MA, Amorim MAL. A farmacovigilância e sua importância no monitoramento das reações adversas a medicamentos. *Revista Saúde e Desenvolvimento* 2013; 4:33-56.
8. Harpaz R, Vilar S, DuMouchel W, Salmasian H, Haerian K, Shah NH, et al. Combining signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013; 20:413-9.
9. Hauben M, Bate A. Data mining in drug safety: side effects of drugs essay. *Side Effects of Drugs Annual* 2007; 29:xxxiii-xlvi.
10. Hotho A, Nürnberger A, Paaß G. A brief survey of text mining. <https://pdfs.semanticscholar.org/9643/0cc91ed91fd2d4042fa6fcb7ecf4005d77a7.pdf> (acessado em Set/2018).
11. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457:1012-4.
12. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009; 49:1557-64.
13. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. *PLoS Negl Trop Dis* 2014; 8:e2713.
14. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS One* 2011; 6:e19467.
15. Lamos V, Cristianini N. Nowcasting events from the social web with statistical learning. *ACM Trans Intell Syst Technol* 2012; 3:72.
16. First Workshop on Computational Methods in Pharmacovigilance held during the Medical Informatics in Europe (MIE) Conference, Pisa, Italy, 29 August 2012. *Drug Saf* 2012; 35:1191-200.

17. Gurulingappa H, Rajput AM, Toldo L. Extraction of adverse drug effects from medical case reports. *J Biomed Semantics* 2012; 3:15.
18. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 2012; 91:1010-21.
19. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In: *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*. Uppsala: Association for Computational Linguistics; 2010. p. 117-25.
20. Yates A, Goharian N. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In: Serdyukov P, Braslavski P, Kuznetsov SO, Kamps J, Rüger S, Segalovich EA, et al., editors. *Advances in information retrieval*. Berlin: Springer; 2013. p. 816-9.
21. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010; 6:343.
22. Ginn R, Pimpalkhute P, Nikfarjam A, Patki A, O'Connor K, Sarker A, et al. Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark. <http://www.nactem.ac.uk/biotxtm2014/papers/Ginnetal.pdf> (acessado em 01/Out/2018).
23. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf* 2014; 37:343-50.
24. Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. In: *SHB'12 Proceedings of the 2012 International Workshop on Smart Health and Wellbeing*. <https://dl.acm.org/citation.cfm?id=2389713> (acessado em 01/Out/2018).
25. Portal Sinan. Calendário epidemiológico. <http://portalsinan.saude.gov.br/calendario-epidemiologico> (acessado em Set/2018).
26. Duval F, Caffarena E, Cruz O, Silva F. Mining for adverse drug events on twitter. In: *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*. <http://www.scitepress.org/PublicationsDetail.aspx?ID=hxPWwh5Sjzw=&t=1> (acessado em 01/Out/2018).
27. Cai M-C, Xu Q, Pan Y-J, Pan W, Ji N, Li Y-B, et al. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic Acids Res* 2015; 43:D907-13.
28. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. *Nat Lang Eng* 1995; 1:83-108.
29. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010; 17:507-13.
30. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: *Proceedings of the AMIA Symposium*. Bethesda: National Center for Biotechnology Information, U.S. National Library of Medicine; 2001. p. 17-21.
31. Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* 2009; 18:1176-84.
32. Dias P, Ribeiro CF, Marques FB. Medidas de desproporcionalidade na detecção de sinal em farmacovigilância. *Revista Portuguesa de Farmacoterapia* 2014; 6:28-32.
33. van Puijebroek EP, Diemont WL, van Grootheest K. Application of quantitative signal detection in the Dutch spontaneous reporting system for adverse drug reactions. *Drug Saf* 2003; 26:293-301.
34. EudraVigilance Expert Working Group. Guideline on the use of statistical signal detection methods in the Eudravigilance data analysis system. London: European Medicines Agency; 2006.
35. Evans S, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; 10:483-6.
36. Rothman K, Greenland S. Introduction to categorical statistics. In: Rothman K, Greenland S, editors. *Modern epidemiology*. 2nd Ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 231-52.
37. Klarreich E. In search of bayesian inference. *Commun ACM* 2014; 58:21-4.
38. Valentín S, Morales A, Sánchez JL, Rivera A. Safety and efficacy of doxycycline in the treatment of rosacea. *Clin Cosmet Investig Dermatol* 2009; 2:129-40.
39. El Emam K, Samet S, Arbuckle L, Tamblin R, Earle C, Kantarcioglu M. A secure distributed logistic regression protocol for the detection of rare adverse drug events. *J Am Med Inform Assoc* 2013; 20:453-61.

Abstract

During the post-marketing period, when medicines are used by large population contingents and for longer periods, unexpected adverse events (AE) can occur, potentially altering the drug's risk-benefit ratio enough to demand regulatory action. AE are health problems that can occur during treatment with a pharmaceutical product, which in the drug's post-marketing period can require a significant increase in health care and result in unnecessary and often fatal harm to patients. Therefore, a key objective for the health system is to identify AE as soon as possible in the post-marketing period. Some countries have pharmacovigilance systems responsible for collecting voluntary reports of post-marketing AE, but studies have shown that social networks can be used to obtain more and faster reports. The current project's main objective is to build a totally automated system using Twitter as a source to detect both new and previously known AE and conduct the statistical analysis of the resulting data. A system was thus built to collect, process, analyze, and assess tweets in search of AE, comparing them to U.S. Food and Drug Administration (FDA) data and the reference standard. The results allowed detecting new and existing AE related to the drug doxycycline, showing that Twitter can be useful in pharmacovigilance when employed jointly with other data sources.

Drug and Narcotic Control; Biological Ontologie; Natural Language Processing; Social Media; Database

Resumen

Durante el período de poscomercialización, cuando grandes poblaciones consumen medicamentos durante períodos más prolongados de tiempo, se pueden producir eventos adversos (EA) inesperados, lo que puede alterar la relación riesgo-beneficio de los medicamentos. Esta situación es suficiente para exigir una acción regulatoria. Los EA son agravios a la salud que pueden surgir durante el tratamiento con un producto farmacéutico, los cuales, durante el período de poscomercialización del medicamento, pueden requerir un aumento significativo de cuidados de salud y resultar en lesiones innecesarias para los pacientes, muchas veces fatales. Por lo tanto, el hallazgo anticipado de EA durante el período de poscomercialización es un objetivo primordial del sistema de salud. Algunos países cuentan con sistemas de vigilancia farmacológica, responsables de la recogida de informes voluntarios de EA durante la poscomercialización, pero algunos estudios ya demostraron que, con la utilización de las redes sociales, se puede conseguir un número de informes mayor y más rápido. El objetivo principal de este proyecto es construir un sistema totalmente automatizado que utilice Twitter como fuente para encontrar nuevos EA y ya conocidos, además de realizar un análisis estadístico de los datos obtenidos. Para tal fin, se construyó un sistema que recoge, procesa, analiza y evalúa tweets en búsqueda de eventos adversos, comparándolos con datos de la Agencia Americana de Control de Alimentos y Medicamentos (FDA) y del estándar de referencia construido. En los resultados obtenidos, conseguimos encontrar nuevos eventos adversos y ya existentes, relacionados con el medicamento doxiciclina, lo que demuestra que Twitter, cuando es utilizado junto a otras fuentes de datos, puede ser útil para la farmacovigilancia.

Control de Medicamentos y Narcóticos; Ontologías Biológicas; Procesamiento de Lenguaje Natural; Medios de Comunicación Sociales; Base de Datos

Recebido em 24/Fev/2017
Versão final reapresentada em 22/Set/2018
Aprovado em 18/Out/2018