* + 1. Dengue Tweet Analysis
    2. *Introduction*

Dengue, also known as “breakbone fever” for the severe myalgia and joint pain experienced by patients, is a major cause of morbidity and mortality around the world. It is caused by a Flavivirus that is transmitted to humans when they are bitten by infected mosquitoes. There are four distinct serotypes of the dengue virus (DEN 1–4), all of which cause disease in humans that ranges from asymptomatic infection to severe, fatal hemorrhagic illness. Recovery from infection provides serotype-specific immunity but does not protect frominfection with other serotypes of the dengue virus.

* + 1. Primary symptoms of dengue appear three to 15 days after the mosquito bite and include highfever and severeheadache, with severe pain behind the eyes that is apparent when trying to move the eyes. Other associated symptoms are joint pain, muscle and bone pain,rash, and mild bleeding. Many affected people complain of low back pain. The lymph nodes of the neck and groin may be swollen. Young children and people infected for the first time typically have milder symptoms than older children and adults.
    2. Dengue hemorrhagic fever starts with the typical signs and symptoms of dengue as described above. The fever lasts from two to seven days. After the fever begins to abate, symptoms occur that are related to increased permeability of the capillary blood vessels. These symptoms can include severe abdominal pain, prolonged vomiting, and breathing problems. Bleeding tendencies, including easy bruising, nosebleeds, bleeding gums, skin hemorrhages, and even internal bleeding may occur. The disease may progress to failure of the circulatory system, leading to shockand death.
    3. With no effective drug therapy or vaccine, control of the mosquito vector and surveillance for clinical infections are the primary public health tools available to fight dengue. Early identification and location of outbreaks can help target intervention campaigns to reduce existing mosquito populations and breeding areas in high-risk locations. The goal of such campaigns is to minimize the spread and impact of an outbreak, but to be effective, intervention needs to start as soon as possible.
    4. Social media, such as the microblogging platform Twitter, provides digitized data continuously 24 hours per day. Posts on Twitter, or tweets, are limited to 140 characters orless, and users tweet to update friends on their activities and thoughts. The content of tweets, therefore, varies wildly—from social commentary to what the user is having for dinner. Most tweets are publicly available through the Twitter application programming interface (API). A pseudo-random sample of tweets meeting user-specified criteria can be obtained relatively easily and free of cost from the Twitter API. Twitter is also heavily used in many resource-limited areas where other sources for electronic disease surveillance are limited.
    5. Multiple investigators have mined tweets for information about the behaviors, moods, and habits of Twitter users, and some have also looked for information to inform disease surveillance. Investigators used Twitter to monitor influenza activity in the United States during the H1N1 pandemic in 2009–2010 and noted good correlation with the number of new influenza cases as collected by public health authorities. Similarly, Collier et al. found a moderately strong association between World Health Organization/National Respiratory and Enteric Virus Surveillance System laboratory incidence data for influenza and the incidence of tweets mentioning influenza during the 2009–2010 influenza season in the United States. Outside of the United States, Chunara et al. compared the volume of cholera reports for Haiti collected from HealthMap (http://www.health-map.org) and Twitter posts with the number of new cholera cases collected via standard surveillance methods by the Haitian Ministry of Public Health. They found a statistically significant positive correlation between the combined HealthMap/Twitter data and the incidence of cholera as collected by the Haitian Ministry of Public Health data (Pearson correlation coefficients ranging from 0.76 to 0.86). Another study by Chan et al. found significant positive correlations (Pearson correlation coefficients from 0.82 to 0.99) between the number of tweets mentioning “dengue” or similar phrases and dengue incidence as measured by public health authorities in Bolivia, Brazil, India, Indonesia, and Singapore.
    6. If a subset of tweets that mimics the true incidence (i.e., the count of new cases) of a disease in a population could be reliably identified, it would be relatively simple to set up a continuous feed of tweets from the Twitter API, process the raw tweets to extract the appropriate tweet subset, and feed those tweets directly into an electronic disease surveillance application. This would provide an inexpensive, yet timely, surrogate disease surveillance data source.

# Tweet Collection

Tweets are collected using Version 1.0 of the free Twitter public API, which allows an individual to request a feed of public tweets matching specific search criteria. Each request, or query, returns a 1% pseudo-random sample of all tweets meeting those criteria, although then precise tweet selection process used by the API has not been disclosed by Twitter.

Here's a list of different twitter fields and their occurrences in a 50,000 tweet dataset:

id 50000

id\_str 50000

created\_at 50000

text 50000

source 50000

truncated 50000

in\_reply\_to\_status\_id 50000

in\_reply\_to\_status\_id\_str 50000

in\_reply\_to\_user\_id 50000

in\_reply\_to\_user\_id\_str 50000

in\_reply\_to\_screen\_name 50000

user.id 50000

user.id\_str 50000

user.name 50000

user.screen\_name 50000

user.location 50000

user.url 50000

user.description 50000

user.protected 50000

user.followers\_count 50000

user.friends\_count 50000

user.listed\_count 50000

user.created\_at 50000

user.favourites\_count 50000

user.utc\_offset 50000

user.time\_zone 50000

user.geo\_enabled 50000

user.verified 50000

user.statuses\_count 50000

user.lang 50000

user.contributors\_enabled 50000

user.is\_translator 50000

user.profile\_background\_color 50000

user.profile\_background\_image\_url 50000

user.profile\_background\_image\_url\_https 50000

user.profile\_background\_tile 50000

user.profile\_image\_url 50000

user.profile\_image\_url\_https 50000

user.profile\_banner\_url 41614

user.profile\_link\_color 50000

user.profile\_sidebar\_border\_color 50000

user.profile\_sidebar\_fill\_color 50000

user.profile\_text\_color 50000

user.profile\_use\_background\_image 50000

user.default\_profile 50000

user.default\_profile\_image 50000

user.following 50000

user.follow\_request\_sent 50000

user.notifications 50000

geo 48656

coordinates 48656

place 48664

contributors 50000

retweet\_count 50000

favorite\_count 50000

favorited 50000

retweeted 50000

lang 36041

entities.hashtags[].text 11024

entities.hashtags[].indices[] 11024

entities.user\_mentions[].screen\_name 22368

entities.user\_mentions[].name 22368

entities.user\_mentions[].id 22368

entities.user\_mentions[].id\_str 22368

entities.user\_mentions[].indices[] 22368

entities.urls[].url 3662

entities.urls[].expanded\_url 3662

entities.urls[].display\_url 3662

entities.urls[].indices[] 3662

possibly\_sensitive 5339

entities.media[].id 1736

entities.media[].id\_str 1736

entities.media[].indices[] 1736

entities.media[].media\_url 1736

entities.media[].media\_url\_https 1736

entities.media[].url 1736

entities.media[].display\_url 1736

entities.media[].expanded\_url 1736

entities.media[].type 1736

entities.media[].sizes.thumb.w 1736

entities.media[].sizes.thumb.h 1736

entities.media[].sizes.thumb.resize 1736

entities.media[].sizes.large.w 1736

entities.media[].sizes.large.h 1736

entities.media[].sizes.large.resize 1736

entities.media[].sizes.small.w 1736

entities.media[].sizes.small.h 1736

entities.media[].sizes.small.resize 1736

entities.media[].sizes.medium.w 1736

entities.media[].sizes.medium.h 1736

entities.media[].sizes.medium.resize 1736

geo.type 1344

geo.coordinates[] 1344

coordinates.type 1344

coordinates.coordinates[] 1344

place.id 1336

place.url 1336

place.place\_type 1336

place.name 1336

place.full\_name 1336

place.country\_code 1336

place.country 1336

place.bounding\_box.type 1336

place.bounding\_box.coordinates[][][] 1336

entities.media[].source\_status\_id 621

entities.media[].source\_status\_id\_str 621

place.attributes.street\_address 2

List Of Symptoms

A symptom is any subjective evidence of disease, while a sign is any objective evidence of disease. Health-care professionals use symptoms and signs as clues that can help determine the most likely diagnosis when illness is present. Symptoms and signs are also used to compose a listing of the possible diagnoses. This listing is referred to as the differential diagnosis. The differential diagnosis is the basis from which initial tests are ordered to narrow the possible diagnostic options and choose initial treatments.







