Table of Contents

## Methods

### Norwegian Data

We used weekly number of primary care consultations for influenza (reported using ICPC-2 code R80) from week 1, 2006 through to week 40, 2015, for each of the 19 counties in Norway.

### South Sudanese

blah, 5 health facilities

### Statistical methods to determine 'extreme periods'

We were primarily interested in comparing three statistical methods:

* Moving epidemic method
* Mean ± standard deviation method
* Linear regression method

The moving epidemic method is widely used in the influensa surveillance world...

The mean ± standard deviation method is a well-known technique. A number of historically comparable weeks are chosen as the baseline, their means and standard deviations are calculated, and prediction intervals are generated. If the new observation is higher than a predetermined cutoff, the new observation is determined to be extreme. For our purposes, we considered historical weeks corresponding to the same week number of interest as the baseline, and chose cutoffs corresponding to two-sided alpha of 0.1, 0.05, and 0.01.

The linear regression method fits a linear regression model to the previous 8 weeks of data, generates prediction intervals for the current week, and if the new observation is higher than a predetermined cutoff, the new observation is determined to be extreme. For our purposes, we chosecutoffs corresponding to two-sided alpha of 0.1, 0.05, and 0.01.

#### Norway

For each county, the moving epidemic method was run three times. Firstly, it used all available data to define extreme periods (defined as medium or higher intensity) of influenza activity -- this was taken to be the gold standard and used as the baseline for comparisons. Secondly, it was run in a retrospective manner, for each year using all data prior to that year. Thirdly, it was run using two years of prior data for each year.

For each county, the mean ± standard deviation method was run twice. Firstly, it was run in a retrospective manner, for each year using all data prior to that year (up to a maximum of 5 years). Secondly, it was run using two years of prior data for each year.

For each county, the linear regression method was run once, using 8 weeks of historical information.

Finally, we also applied a method where weeks were randomly assigned to extreme/not-extreme (using a frequency distribution taken from the results of the moving epidemic model with all data), to examine the how much each method surpassed random assignment.

#### South Sudan

For each county, the moving epidemic method was run two times. Firstly, it used all available data to define extreme periods (defined as medium or higher intensity) of influensa activity -- this was taken to be the gold standard and used as the baseline for comparisons. Secondly, it was run using two years of prior data for each year.

For each county, the mean ± standard deviation method was run using two years of prior data for each year.

For each county, the linear regression method was run once, using 8 weeks of historical information.

Finally, we also applied a method where weeks were randomly assigned to extreme/not-extreme (using a frequency distribution taken from the results of the moving epidemic model with all data), to examine the how much each method surpassed random assignment.

### Comparison of methods

To allow the methods to be evaluated over the same time periods, the first two years of results were removed for each place and method. For each place, method, and year, the number of "strikes" (i.e. weeks above threshold) were counted in a cumulative manner. Spearmans correlation coefficient was then calculated for eache place and method, comparing the cumulative number of strikes for each method against the gold standard.

We then took each country/method combination and performed unpaired t-tests to compare the mean of the correlation coefficients of the methods of interest against the correlation coefficients of the randomly generated strikes. For South Sudan, we noticed that the correlation coefficients for Doro were remarkably lower than the others, so in a posthoc decision we ran an additional set of t-tests excluding Doro.

## Results

### Influenza in Norway

* The methods using only two years of data performed worse than the methods using all years available.
* The mean ± standard deviation method performed worse than the MeM methods, however, this might be merely due to more statistical similarity between the MeM methods, as our baseline is not a true gold standard (as there is no "truth" in this analysis)
* From a visual observation, the linear regression algorithm performed approximately as well as the mean ± standard deviation method, even though it only used 8 weeks of historical data

### Malaria in South Sudan

* The mean ± standard deviation method performed worse than the MeM methods, however, this might be merely due to more statistical similarity between the MeM methods, as our baseline is not a true gold standard (as there is no "truth" in this analysis)
* From a visual observation, the linear regression algorithm performed worse than the mean ± standard deviation method
* From a visual observation, using alpha=0.01 appeared to give the best results for linear regression
* Due to the outlier Doro, the linear regression algorithm did not perform significantly better than random, however, after removing Doro, linear regression at alpha=0.01 did perform significantly better than random

