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

#### 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 influensa 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, 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, to examine the how much each method surpassed random assignment.

### Comparison of methods

For each method and year, the number of "strikes" (i.e. weeks above threshold) were counted in a cumulative manner. Spearmans correlation coefficient was then calculated, comparing the cumulative number of strikes for each method against the gold standard.

## Results

* 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)
* The linear regression algorithm performed approximately as well as the mean ± standard deviation method, even though it only used 8 weeks of historical data



