*Team members*:

Fred Lu,

Shihao Yang,

John Brownstein,

AthenaResearch team, athenahealth,

Mauricio Santillana (Principal Investigator 1),

Samuel Kou (Principal Investigator 2)

*Data sources*:

1. Google search frequencies (From Google Trends and Google Health API)

2. Electronic Health Records data from Athenahealth.

*Description of methodology:*

Our method combines information from electronic health records (EHR) and Internet users’ search activity (Google) with historical flu activity to produce flu forecasts for the United States. Our methodology is based on a regularized multivariate regression model that dynamically selects the most appropriate variables for optimal flu estimates and automatically identifies which variables are important for prediction in every week. An early version of this methodology can be found in:

 S. Yang, M. Santillana, and S. C. Kou. Accurate estimation of influenza epidemics using Google search data via ARGO. Proceedings of the National Academy of Sciences, 112.47, pp: 14473-14478, 2015.