

Estimating FluNearYou Correlation to ILINet at Different Levels of Participation

Eric V. Bakota*, Eunice R. Santos and Raouf R. Arafat

Health Department, City of Houston, Houston, TX, USA

Objective

Our objective is to provide evidence for the data quality of Flu Near You (FNY) by evaluating the national and Houston datasets against CDC ILI data.

Introduction

Flu Near You allows individuals to volunteer to be a sentinel node of the syndromic surveillance (SyS) network. The platform has the potential to provide insight into the spread of influenza-like illness (ILI). CDC's ILINet is the gold standard for tracking ILI at the national level, but does not track into the local level. Local health departments (LHD) frequently express a need for granular data specific to their jurisdictions. FNY attempts to meet this need by collecting and sharing data at the zip code level. Knowing how well FNY data correlates to ILINet data will give local health departments an important tool to communicate the arrival of influenza to their jurisdiction. However, there is significant skepticism at the quality of FNY data as compared to validated datasets.

Methods

FNY pushes out a weekly survey to each user. The survey tracks if and when a user (and his or her family) has received a flu shot and experienced ILI. The data were deidentified and provided by the Skoll Global Threats Fund to the Houston Health Department (HHD). The FNY data were compared to ILINet's national summary of influenza-like illness and influenza positive tests by estimating the correlation coefficient for the 2014-2015 influenza season. FNY total ILI counts were correlated to total positive influenza tests and FNY percent ILI was compared to ILINet's unweighted percent ILI. The mean correlation coefficients were estimated by bootstrapping FNY data ($n = 1000$ at each stratum). Mean correlation coefficients over 1000 bootstraps were estimated for a sequence of weekly participation rates from $N=10$ to $N= 10435$ in increments of 10. Bootstrapped samples were stratified by ZIP code to account for fluctuations in weekly participation for both FNY and ILINet, as both datasets see an increase in user participation during influenza season. R version 3.2 was used for all analyses; HHD received the line-list dataset from FNY that contained nearly 400,000 entries. Each entry corresponds to a single person (either the user or the user's family member) and his or her symptoms for the preceding week. FNY is voluntary and not all users contribute each week; with the data being deidentified there is no way to connect entries to user profiles or user families. As such, each entry is treated as independent. ILINet data came from CDC.

Results

Key Finding:

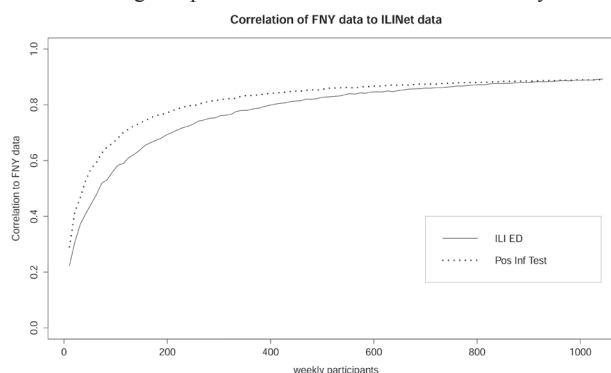
- Correlation of the full FNY dataset against ILI is very high ($r^2 = .94$).
- Correlation of the full FNY dataset against positive influenza tests is also high ($r^2 = .92$).
- Weekly reports from < 200 weekly users have high variance in their correlation to ILINet and a moderate correlation coefficient ($r^2: 0.3 - 0.7$).

• Reports from ≥ 200 weekly users seems to be the inflection point for diminishing returns with respect to improving mean correlational coefficient of the pseudo-replicated data.

- FNY correlates well with both ILI visits as well as positive influenza tests.
- At low participation counts, (< 400 per week) FNY correlates better with positive influenza tests than percentage with ILI.
- Overall, FNY data correlates well with national ILINet data, even at limited participation levels.

Conclusions

Approximately two-thirds of the counties within the United States have a population of $< 50,000$. As such, FNY provides a simple, low-cost opportunity for public health officials within those jurisdictions to obtain data that reasonably mirrors ILINet. For larger jurisdictions, FNY is another tool available to track and identify seasonal influenza and engage the public on prevention. This comparison supports the idea that FNY will give local officials in smaller jurisdictions more confidence to guide public health action in their community.



Keywords

Influenza; Flu Near You; Local Data

Acknowledgments

I would like to thank Skoll Global Threats Fund for sharing the Flu Near You data.

*Eric V. Bakota

E-mail: eric.bakota@gmail.com

