Geocoding

Overview: FluSurv-NET performs geocoding and linkage to US Census data using standardized methods for all cases identified as part of surveillance activity. Geocoding at the census tract and residential address level has the potential to help us understand socioeconomic predictors, describe health disparities and identify outbreaks in long-term care facilities.

Protocols, CRF, Database: Geocoding is done now automatically through Accumail for address verification and ArcGIS, mostly within the EIP database. Approximately 80% of flu cases are geocoded to residential address. The remaining cases, which are generally P.O. boxes, rural routes, directional addresses (e.g. 1.5 miles past mile marker 55) are manually reviewed by the NMDOH geocoder in the past. In recent years, this person has not had the bandwidth to review these addresses, so they are usually left as is. Geocoding has the potential to help us understand neighborhood factors at the census-tract level and identify outbreaks in long-term care facilities.

When it is time to geocode, a number of steps should be taken:

1. Ask the EIP IT contact (Sri HarshaChinta) to batch geocode all addresses, and to run the facility address match. This cleans up any geocoding that was missed, assigned the correct addtype to residential addresses, and compares addresses to the addresses of known facilities such as LTCFs, correctional facilities, homeless services, etc. That way, the correct addtype can be assigned based on the address.
2. Then, pull all the geocoding data necessary from EIP via SQL.
    a. The SQL code can usually be found in the data file from the prior season labeled "preliminary_pull" in the file name.
3. All addresses that were not geocoded, or have an addtype of 8 need to be checked. There are two options for this:
    a. An R script was created that automatically returns the first search result in Bing. Up to 150 addresses at a time can be run through this. The output spreadsheet has the first search result, so usually you can see whether this is a residential address (key words like single family, apartment, condo, trailer, etc). Any addresses where the search result is unclear should be looked up manually. Using this method will cut down on the amount of time spent manually searching for addresses.
        i. R script found here: K:\FluSurv 2018-2019\18-19 Geocoding\Address Search
    b. Alternate: assign someone to manually search these addresses to see if there is a misspelling or something wrong with the address, and to determine the correct addtype.
    c. For all addresses that have misspellings or need to be updated: these should be updated in the EIP database, where hopefully the correct address can be automatically geocoded.
4. Once all remaining addresses have been checked, the dataset can be pulled again from SQL.
5. This SQL dataset needs to be merged with the REDCap dataset, in order to add the variable NHTYPE (where did the patient reside at the time of admission?). This variable helps to assign the correct addtype.
6. The person completing geocoding will then need to use the table in the document below to assign the correct addtype.
    a. K:\COVID NET 2020\GEOCODING\Jan 2021 - Sept 2021\ COVIDNET\ NM Geocoding instructions 05132021
7. Once this is done, the dataset is ready to clean up (remove unnecessary variables) and transmit to CDC via SAMS