

# Demographic Health Analysis by Incorporation of Census Data with Patient Records

Christopher R. Cuellar\* and Wayne Loschen

Johns Hopkins University Applied Physics Laboratory, Columbia, MD, USA

## Objective

The objective of this project is to enable a deeper analysis of patient health by correlating patient health records with the census demographic data. Based upon these correlations, the ESSENCE system will be enhanced with new query filtering capabilities.

## Introduction

Electronic disease surveillance canonically represents analysis performed on health records with respect to their syndromes, complaints, lab data, etc. This data can tell the story of a patient's current status but does not provide a holistic look at the where the patient is from. By incorporating census data, a deeper examination of the patient's area can be performed which may result in discovery of risk factors associated with race, economic status, and culture.

## Methods

Data was gathered from census surveys conducted by the US government in a comma delimited file [1]. Datasets used included household income, poverty status in the past 10 years, and race. All data within each file was associated with a census ZCTA block. ZCTA's, while similar to zip codes, do not share a one-to-one correlation with zip codes, and subsequently, need to be mapped appropriately to use them with zip code-based health records. The most common approach to this mapping is to calculate the centroid of the ZCTA region and project that onto the zip code region. We used a dataset that performed such a mapping for our zip code to ZCTA correlation [2]. Finally, through collaboration with the Florida Department of Health, we developed several groupings for the data. Partitions of health records were then built based upon factors such as percentage of race or ethnicity in a specific ZCTA, median household income, and predominant race for the resident ZCTA block. These groupings were then incorporated into the Query Portal feature of the system.

## Results

Users can now filter for all health data by a demographic factor, such as finding all patients from a predominantly Hispanic zip code, or all fever cases from zip codes with a median income level of \$50,000 or more. Primary complications were due to the usage of ZCTA for census data and the usage of zip code for patient records. Since there is no exact one-to-one mapping, it is impossible to assure that patients reside within a specific ZCTA without a more granular location being specified. ZCTA data is also subject to the number of people responding to surveys within a given area and can be unreliable for areas of low participation. The binning for census data is variable depending on the location which it is being applied, there is no perfect separation that can be applied especially when considering median household income.

## Conclusions

Application of census data can be burdensome, especially as zip codes have a tendency to change in shape; however, the usefulness of census data in determining societal risk factors can be immeasurable. Future work includes developing the capability to automatically

determine changes in ZCTA and zip code representations. Development of appropriate binning for census data is a problem of localization as features such as median household income can be overall higher or lower within a state compared to other states. Finally, inclusion of a metric comparing cost of living should be added to get a better idea of the worth of specific median household incomes. This talk will detail the technical approaches, including complications, along with the types of data incorporated and the resulting forms of analysis possible. The datasets themselves will be discussed in terms of their granularity and what information they can provide.

## Keywords

Census Data; System Development; Query Improvement

## References

- 1) U.S. Census Bureau; using American FactFinder; [http://factfinder2.census.gov](http://factfinder2.census.gov;); (May 2014)
- 2) UDS Mapper; <http://www.udsmapper.org>

\*Christopher R. Cuellar

E-mail: [Christopher.Cuellar@jhuapl.edu](mailto:Christopher.Cuellar@jhuapl.edu)

