**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 12/05/16 |
| **Project Title** | **Decentralized and HIPPA Compliant Geocoding to Characterize Community and Environmental Exposures for Multi-Site Studies** |
| **Tentative Lead Investigator (first author)** | Cole Brokamp |
| **Tentative Senior Author (last author)** | Patrick Ryan |
| **All other authors**  | TBD |
| **Sites Involved** | **ALL** (eMERGEIII Geocoding Supplement Working Group) |
| **Background / Significance** | Geocoding (assigning latitude and longitude coordinates to postal addresses) is a necessary first step to characterize study participants’ community and environmental characteristics in the vicinity of their residence, schools, etcGeocoding involves the use of identifying information (addresses) and therefore must be conducted in a HIPPA and IRB compliant mannerGeocoding and assigning environmental / community characteristics for participants in multi-site studies governed by multiple IRBs presents logistical challenges and, typically, requires:1. Specific IRB approval for a central site to conduct the geocoding / analyses OR
2. A decentralized approach whereby each site conducts geocoding independently, links geocoded locations to the desired geographic data, and then separate the geographic data from the addresses to create a de-identified dataset for analyses.
* Approaches (1) and (2) have significant limitations including IRB approvals, inconsistent geocoding results, expertise available at each study site, and others

Therefore, we propose a novel, decentralized approach whereby one site (Cincinnati) develops standalone geocoding software capable of producing geocodes and deriving community and environmental exposures. This software will allow each study site to independently geocode their sites addresses and link community / environmental exposures to these. Removal of the addresses from this dataset at each site will produce a non-identifiable (and therefore IRB compliant) dataset of community and environmental exposures to be merged by the eMERGE coordinating center.  |
| **Outline of Project** | * Execute a decentralized approach to extract census block FIPS identifier and amount of nearby greenspace by the following process:
	+ Specialize our standalone geocoding software to produce census block FIPS and a satellite based greenspace measure within a 400 meter radius
	+ Supply the software to participating eMERGE sites and have them locally analyze their cohorts and strip PHI
	+ Each site will then provide us with the patient IDs, FIPS identifier, and greenspace estimate as well as the confidence score and geocoding method details
	+ Note that each site will not share any PHI and the software will be run locally without any connection to the internet
* Summarize these geocoding and exposure estimate results
* Obtain feedback from individual sites on how usable they found the software
 |
| **Desired****Variables** **(essential for analysis****indicated by \*)** | N/A |
| **Desired Data** | Each site will locally apply our geocoding software and return only non PHI data, which will include the patient ID and environmental exposure estimates generated by our software |
| **Planned Statistical Analyses** | Summarize precision of geocoding results and how successful the decentralized approach is. |
| **Ethical considerations** | none |
| **Target Journal** | Journal of Exposure Science and Environmental Epidemiology |
| **Milestones\*\*** | **Gain Project Approval and Identify Participating Sites:** February 2017**Complete Specialized Software:** April 2017**Participating Sites Apply Software to Cohorts:** May - June 2017**Summarize Results and Prepare First Draft of Paper:**  July 2017**Complete Second Draft of Paper and Submission to Journal:** June 2017 |

**\*\*** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.