**Data Management Plan.**

***Data Collection and Generation.*** Data collected will mostly pertain to those asked in nationally conducted surveys conducted by CDC/NCHS (e.g., NHIS, NHANES or BRFSS), surveillance sources such as NCHHSTP AtlasPlus, administrative datasets which include patient demographics, encounters, diagnoses, procedures, labs, medications, vaccinations, and other previously collected quantitative/qualitative data collected / generated as part of patient care. As part of this project, data related to laws and policies at various geographies (e.g., state, county, city, or county) will also be compiled. ZCTA/ZIP-level SES indicators from the US Census may be combined with aggregated patient data by residential zip code. Some of these data will become public health data, digitally recorded factual material that can be used as a basis for public health findings, conclusions, and implementation. Public health data includes research and non-research data.

***Data Standards.*** The source data will be extracted, transformed, and loaded (ETL) into the Urban Health Collaborative HIPAA Secure Database environment using appropriate ETL tools available to in a secure environment. MySQL, is a cloud-based scalable database management system that will enable us to securely record, update and manage data in compliance with GDPR, PCI, HIPPA and various regulatory standards. In terms of data interoperability we will adhere to US census Geographic identifiers coding for spatial linkages.

***Data Access.*** All data will be stored on a HIPAA compliant, password protected server. Access will be restricted to Applicant Team members with a secure access account and a valid phone number or mobile device with the Duo Two-Factor Authentication application. Two-factor authentication strengthens security and protects against phishing, social engineering, and password brute-force attacks. Our team has access to the Health Informatics lab, HIPAA compliant and de-identified database servers and webserver, and dedicated CPH IT support for creating a data pipeline for using Power BI, Tableau, and other business intelligence and analytics software for dashboards. Only de-identified (i.e., sensitive identifiable or potentially identifiable information are removed) data will be available for public use. De-identified data and an accompanying data dictionary, codes, and other documentation relevant to use of the data set will be deposited in a sustainable repository. Data that cannot be de-identified can be provided on request under a data use agreement (DUA) that covers security protocols, data protection, privacy, confidentiality, intellectual property, and all other rights associated with the data. Where and when necessary, obfuscation, masking, and other de-identification best practices will be used to ensure complete de-identification of data and restrict attempts at re-identification. For data underlying scientific publications, corresponding data will

be made available unless already available via a release or sharing mechanism. At minimum data will consist of a machine-readable version of the data tables shown in the paper. Analysis code will be made available on a online version control platform such as GitHub; public-access data associated with analyses will be distributed either via GitHub or via cloud storage.

***Metadata.*** Data information (i.e., data dictionaries, entity relationship diagrams, data models, and data lineage information) will be collected and documented by the health informatics team in a standardized way. All data released will have documentation to describe the method of data collection, what the data represents, and potential limitations for secondary use of the data. We will document our research data and metadata with the Data Documentation Initiative (DDI) standards; importantly with DDI we are able to represented our relational MySQL database in a standard format that is interoperable most large data providers such as the Inter-university Consortium for Political and Social Research (ICPSR).

***Data archiving and preservation.*** When sending data to the CDC this will not be required. If long term data preservation and data stewardship is required, we will share data via a data archive organization such as the Inter-university Consortium for Political and Social Research (ICPSR). Since our metadata and data is organized via DDI, our data assets should be fully compliant with ICPSR submission requirements.

***Plans for quarterly data sharing with funding agency.*** As required by the CDC and/or the coordinating center all relevant public health data requested will be shared quarterly with the funding agency through the appropriate requested method such as secure file transfer protocol, Github repository, virtual private network, dashboard, or other method as instructed.