

Geography and Epidemiology of Health Data for the City of Camden: Creating a Neighborhood Health Indicator Database

Investigators

Co-principal investigators

- Jeffrey Brenner, MD - Cooper Health System, Department of Family Medicine
- Brigette M. Baumann, MD - Cooper Health System Department of Emergency Medicine
- Derek Ziegler, Program Manager - CAMConnect

Other investigators

- Ian Hoffman, MS4, Health Policy Fellow - Cooper Health System, Department of Family Medicine

Objectives/Purpose

In recent years cities around the country have taken it upon themselves to construct important advanced information systems with integrated and recurrently updated information on neighborhood conditions. A consortium of organizations by the name of the National Neighborhoods Indicator Partnership (NNIP), part of the Urban Institute, has been instrumental in the scale-up of this technological capacity to evaluate cities on an ongoing basis (Kingsley 1998). Creation of this capacity, which did not exist in any U.S. city a decade ago, represents an important technical and institutional breakthrough.

The City of Camden is fortunate to have membership in the NNIP via CAMConnect, a nonprofit organization dedicated to collecting and disseminating neighborhood data. Their mission is to allow stakeholders to make informed decisions and effective policies to improve the quality of life for Camden citizens. The organization's objectives include:

- 1) Creating and maintaining a standardized, integrated database that contains quality of life indicators for Camden neighborhoods and comparison data for the city, county, state and selected regions; and
- 2) Providing improved access to baseline and change data to stakeholders and the public through reports and online access, in compliance with confidentiality and public access laws.

Previously, the Cooper Health System Department of Family Medicine and CAMConnect have investigated rates of injuries in different Camden neighborhoods by analyzing ED discharge data from Cooper Hospital for 2001-2003 (protocol # 02-058). This expanded research protocol is intended to build on these previous methods with a more comprehensive look at several years of data from all departments at all hospitals in the City of Camden.

Hypotheses

There is likely to be significant variability in rates of different types of health conditions within the City of Camden, as well as significant deviation from national approximations of these health conditions based on surveys of other health surveillance systems. In

addition, there is likely significant correlation between rates of these health conditions and population, housing, and socio-economic variables by census block group, neighborhood, or other geographic area.

Additionally, access to these health data, which are valuable for the community to assess current policy, understand progress, and make decisions on future policy, is an essential part of Camden's attempt towards revitalization. The access of community health data by governmental and community organizations will improve the conditions of life in Camden by guiding these institutions toward more informed public policy with ongoing evaluation.

Background

Within the City of Camden there is a variation in household income, housing values, ethnicity, age, crime, and other demographics between neighborhoods. (website: www.camconnect.org) The ability of residents to advocate for themselves and the ability of the city and local non-profit agencies to target responses to problems is augmented by the availability of neighborhood-level data, as well as smaller defined geographic areas such as census block or point locations.

The City of Camden does not have a health department. The function was taken over by the Camden County Health Department. For this reason, very little health data is available that goes below the county-level much less to the neighborhood, block group, or block level. This project will begin the process of gathering improved neighborhood-level health data. Additionally, morbidity data has often taken the back seat to mortality data in public health. Our proposal will build on both health indicators by improving the currently available mortality data, and making available previously unavailable morbidity data.

Significance of the research

As health care costs continue to rise, acquiring a better understanding of where outreach efforts are needed becomes ever more critical. Understanding the health of a community is a complex task, and requires the proper tools for evaluation of current health status, trends in health, and proposals for future health policy. This study will help to show if Camden neighborhoods are experiencing differences in health outcomes. This study will also show which populations are more susceptible to certain health outcomes. It will also afford groups addressing specific health concerns to monitor their progress and evaluate their programs over time.

Examples of what this database will be used for:

1) The 2000 Uniform Crime Report showed that Camden has the highest murder and violent crime rate of any city in the State of New Jersey. (http://www.njsp.org/info/ucr2000/sec6_ucr2000.htm) This high rate of crime contributes to a high rate of trauma and injuries within the City of Camden. Information about geographical patterns of accidents and injuries will be useful to residents, organizations, health providers, and city leaders. This data will also be useful in implementation of

violence prevention programs run through community organizations and/or governmental initiatives.

2) Hospital discharge data of hospitalizations due to violence will be used to compare other methods of estimating community violence in Camden. These data are available through the New Jersey Department of Health and Senior Services, and assault rates per 100,000 residents per zip code have been calculated by the Violence Institute of New Jersey and UMDNJ.

3) Data from the emergency departments of hospitals can provide important indicators of access to health care and health disparities. One measure these data can address is the issue of hospitalizations for ambulatory care sensitive conditions. If patients seek medical attention from an emergency department instead of a primary care physician, it suggests several things: the patient may not have access to a primary care physician, or the patient may not take the problem seriously until it becomes acute enough to warrant a visit to the emergency room. Studies (Parker and Schoendorf 2000) have shown a correlation between economic status and discharges for ambulatory care sensitive conditions (ACSC) in children. Other studies (Laditka et al 2003) have shown that African-American and Hispanic populations have a higher rate of hospitalizations for ACSC. Creating capacity to investigate the extent of this correlation in Camden City is at the core of this research proposal.

4) Other studies have investigated the difference in rate of hospitalizations for ACSC in metropolitan areas in the United States and Canada. Billings et al (1996) found that “average admission rates in low-income areas were as much as 3.7 times greater than rates in higher-income areas, with individual low-income ZIP code areas having rates more than twenty times higher than those in some of the more affluent ZIP code areas.” Burr et al (2003) found that ICU admissions for ACSC was related to access to primary care. The degree and type of insurance coverage is also correlated to the rate of hospitalization for ACSC (Pappas et al 1997). The expansion of our database to include all health indicators will also expand the discussion of comparison between Camden and other cities.

5) Previous studies have used ED registries to estimate national rates for different types of trauma (Kyran, 1999). Other groups have used registries to look at occupational injuries (Forst, 1999), or to ascertain local information on gun violence (Kellermann, 2001). Small area analysis using the ED registry allows a research group to build local pictures of injuries in the community such as childhood risk factors for severe injury (Pomerantz, 2001). It has been proposed that ED registry data is more accurate than police department records (Shepard, 1989), and therefore is a better estimate of certain types of community injuries such as assaults and gunshots victims. Mortality data on violent deaths has been used for regression with socio-economic, population, and health variables to construct a spatio-temporal history of violence within a community (Wallace, 1998.) Creation of this database will allow researchers to do small-area analysis for any health indicator from motor vehicle crashes, to assaults, or asthma and diabetes.

Study Design

Hospital discharge data for Cooper Hospital, Our Lady of Lourdes Medical Center, and Virtua Health Camden will be requested. In addition, specific patient registries (trauma, perinatology, neonatology, emergency department, radiology, etc.) will be used in cases where the hospital discharge records are insufficient. These routine datasets contain a number of variables, and are often used for internal auditing and research purposes. The requested dataset will contain unique patient identifiers, but will only be maintained with these unique identifiers for purposes of processing the data. Once data has been processed, all unique identifiers will be destroyed. The data requested will contain all discharges based on ICD-9 coding (International Classification of Disease, Ninth Revision), and a reported address within the city of Camden.

In order to better develop our understanding of the underlying patterns of geographic epidemiologic data, we have in the past found it necessary to correlate and expand upon our the discharge data through other hospital data sources. Therefore, from either the databases listed or from periodic chart review, other imperative public health data will be collected. This will include data from fields in the databases, chart review for lab values, extraction from history and physical of charts, or radiology films or notes. Some examples include information from neonatology and perinatology about birth weight or sonographic findings; from the emergency department, type of weapon used and location of assaults; from the trauma unit, location of a motor vehicle crash; from general medicine charts, the hemoglobin A1C of diabetic patients seen in the hospital; from radiology, bone density.

An Excel spreadsheet or SPSS database will be maintained that documents the geographic location data (street-level location, latitude/longitude, census block, census block group, census tract, zip code), discharge data (type of discharge, severity, outcome), time data (year, month, day, hour), and patient demographics (age, sex, race, ethnicity). No patient names or other unique identifiers will be maintained in this research record. Once data has been processed, all unique identifiers will be destroyed. The patient address will be geocoded using an online geocoding service (www.geocode.com). The home address data plotted will not use point location in order to protect the privacy of patients. The smallest aggregate area plotted will be an area with at least 100 residents.

However, in cases where the data collected contains information about the location of the health incident, and this is felt to be an important reportable event, the researchers will publish point location data so long as it does not compromise patient privacy. An example of such an event would be publishing the location of several motor vehicle crashes surrounding a certain intersection.

Additionally, in some cases there will be a need for retrospective chart review to obtain further data. For instance, in the cases of diabetes discharges, this subset will be further analyzed to correlate with medications, primary care physician, or other non-patient identifying items from the chart. Discharges for assault will also be looked at retrospectively to determine local of the incidence, involvement of substances, and relationship to perpetrator.

This data will serve as a reference for comparing Camden City injury data to other external datasets and temporally within the data collected on the City of Camden. The use of this discharge database will allow our group to carry out correlation between health indices and other datasets discussed below.

PHI To Be Obtained

The following chart describes the type of PHI that will be collected, which patients the information will be collected for, sources of the data, and the presentation of that data in the final public dataset. All data will be obtained by researchers listed on the study who have completed HIPAA training, the IRB-approved Human Subjects course, and are registered with the IRB. None of the following will be identified as individual datapoints, rather as aggregates of similar data.

PHI	Source (primary)	Source (secondary)	Diagnoses code limits	Form of data made public*
Pt. Name	EMR	None	All	None
Medical record #	EMR	None	All	None
Date of birth	EMR	None	All	Age (years)
Sex	EMR	None	All	Sex
Address	EMR	None	All	Neighborhood
Race	EMR	None	All	Aggregate race
Date of admission	EMR	None	All	Length of stay
Date of discharge	EMR	None	All	Length of stay
Diagnoses codes	EMR	None	All	Condition categories**
Insurance status	EMR	None	All	Medicare, Medicaid, private, uninsured
Diabetes Medications	EMR	Chart review	Diabetes codes	Diabetes medication class
HgA1C	EMR	Chart review	Diabetes codes	HgA1C
PFT	EMR	Chart review	Asthma, COPD codes	PFT
Asthma, COPD medications	EMR	Chart review	Asthma, COPD codes	Asthma, COPD medication class
Frequency of visits	EMR	Chart review	Asthma, COPD, Diabetes	Frequency of visits
ASD/MI/IHD/CVD* medications	EMR	Chart review	ASD/MI/IHD/CVD	ASD/MI/IHD/CDB medication class
Lipid panel	EMR	Chart review	ASD/MI/IHD/CVD	Lipid panel value ranges
Cardiac stress test	EMR	Chart review	ASD/MI/IHD	Cardiac stress test categorical findings
Birth weight	EMR	Chart review	Birth	Birth weight
Gestational age	EMR	Chart review	Birth	Gestational age
Prenatal/Perinatal medications	EMR	Chart review	Prenatal/perinatal	Prenatal/Perinatal medication class
Ultrasonography categorical findings	EMR	Chart review	Prenatal	Ultrasonography categorical findings
Blood lead levels	EMR	Chart review	All	Lead level ranges
Geographic location of injury	EMR	Chart review	Injuries	Geographic location on map
Assault ROS	EMR	Chart review	Injuries	Assault ROS
Month of admission	EMR	Chart review	Asthma, injuries	Month of admission
Time of day of admission	EMR	Chart review	Injuries	Time of day in blocks of 6 hours

				*All data presented as aggregate unless otherwise noted.
* ASD (atherosclerotic diseases), MI (myocardial infarction), IHD (ischemic heart disease), CVD (cerebral vascular disease)				**Condition categories will be aggregates of multiple diagnoses codes for a certain condition (i.e. asthma, COPD, diabetes, assault, STD, etc.)

These aggregate groups will be presented in tables, charts, and maps that display groups, not individuals. For instance, one of the most detailed potential tables/maps might display number of black (race) male (sex) patients, ages 0-14 (age group), who presented to the ED for falls (diagnoses code), broken down by 6 hour time blocks (time of day), but presented as a yearly figure, with the point where they fell mapped as a point location (geographic location of injury). Nowhere would this data suggest that a particular 4 year-old black male patient fell at a particular point between the hours of 6PM-12AM on a particular day. Rather, the table would show that in a certain neighborhood, a certain number of black male children between the ages of 0-14 were injured in falls during a certain year. The yearly number would also be presented broken down in the 6 hours blocks (time of day) for the group (as a whole, not individual patients), or potentially by month instead of year (if there were significance for monitoring or evaluating a program). Also, one could see on the map where these falls occurred, but that point would not be linked to specific information beyond the aggregate information for the group about age, race, sex, time of day.

PHI Protection Protocol

To ensure that all data remains secure and that all PHI remains confidential, a detailed plan for acquiring, working with, presenting, and destroying the data has been developed.

Original data files and Verified Data Files: All data acquired electronically from various electronic hospital records will be burned to individual CD-ROM discs. The data will then be transferred to a secured storage device that is locked in the department of Family Medicine at Cooper Hospital in the Education and Research Building (E&R). Once data is stored in its original form on this device, the transfer CDs will be destroyed. All data files (named as *original*, *verified*, and *working*) will be password protected and stored on the secured storage device. Only researchers who are partners of this study, who are registered with the Cooper IRB, and who have completed HIPAA training and the online Human Subjects course will have access to these files and passwords, and only on the premises of the E&R building at Cooper Hospital.

The *original data files* will contain all fields of sensitive PHI listed in the proposal and collected from the participating hospitals. The first step of work done with the *original*

data files will be to create a *verified data file* that has been checked for accuracy, duplications, or other problems with the *original data file*. PHI in the *verified data files* that will remain includes all PHI that was in the *original data file*. A final copy of the *verified data file* will be saved to the secured storage device. The *original data files* and the *verified data files* will be stored on the secured storage device for up to 2 years from the date of acquisition for purposes of research methods and quality control. At that time the data will be destroyed, and the storage device will be reformatted to ensure all data is adequately erased.

Neither of these files, the *original data files* nor the *verified data files* will leave Cooper Hospital or be saved on any disk other than the secured storage device in the E&R building at Cooper Hospital.

Working Data Files: Once all data has been verified, PHI such as patient name, medical record number, date of birth, date of admission, date of discharge, home address, and any identifiers of address below the neighborhood (census tract) level will be destroyed or de-identified by aggregation, and from that point on the *working data files* will not contain these sensitive PHI. These files will be the intermediate files that will be used to create the *final public data files*. The process of aggregating data for de-identifying is the purpose of the *working data files*. *Working data files* will be stored on the secured storage device for up to 2 years from the date of acquisition for purposes of research methods and quality control.

Final public data files: The *final data files* that will be publicly published via the department of Family Medicine, periodicals, or CamConnect, will not contain the most sensitive PHI. The PHI in these final public files that will have been destroyed or de-identified include: Patient name, medical record number, date of birth, date of admission, date of discharge, home address, and any identifiers of address below the neighborhood (census tract) level. All other PHI listed will be thus de-identified and available through aggregated groups in tables, charts, or maps that adequately protect the PHI of individuals. Data considered PHI will not be presented as individual cases, only as aggregates.

Coordinating External Datasets:

Population variable data will be used in calculating rates for given geographic areas, spatial analysis, and regression modeling. These datasets include US Bureau of Census files for 1980, 1990, and 2000, on population, housing, and socio-economic variables. Mid-census estimates will be used where appropriate and statistically valid.

CAMConnect, a local non-profit group that acts as data clearinghouse for the city of Camden, will act in close coordination with the Department of Family Medicine on this project. CAMConnect is devoted to creating a learning partnership in Camden, with the belief that sharing and using data will lead to an improved quality of life for all Camden citizens through informed decisions and better policy-making. CAMConnect's valuable datasets will be used as a source for acquiring much of the previously mentioned external data. Data already available through CAMConnect includes census population and housing data up to 2000, socio-economic data for 1980 and 1990, crime arrest data

for the city of Camden for 2000, and other variables on community assets such as churches, community organizations, pharmacies, physicians, and hospitals.

CAMConnect will also have access to finalized datasets and maps created through this research. These data and maps will be disseminated by CAMConnect to community organizations, local action groups, policy advisors, and policy makers for the city of Camden.

Research Plan

Number of Subjects - All patients seen at Cooper Hospital, Our Lady of Lourdes Medical Center, and Virtua Health Camden.

Exclusion and Inclusion criteria - All patients with a valid Camden City address who were seen at Cooper Hospital, Our Lady of Lourdes Medical Center, and Virtua Health Camden during the time period of 1980-2003 will be included in this study.

Recruitment Methods - This is a retrospective study. All patients included in the participating hospital's discharge system will be included.

Informed Consent Process - Given the retrospective nature of this study and the safeguards for retaining subject confidentiality, the investigators request that informed consent be waived. The research involves minimal to no risk to the health or confidentiality of individuals. It utilizes existing data from medical records, and representation of these data will be aggregated to a degree that unique identification of an individual is improbable.

Vulnerable Subjects – This study will include those considered traditionally vulnerable subjects as it will include all residents of the City of Camden seen in a local hospital.

Compensation to Subjects – There will be no compensation to subjects. This is a retrospective study.

Treatment for Research-Related Injuries – Not applicable.

Research Methods and Procedures –

Geocoding:

The geographic location data (street-level data) must be geocoded to obtain useful variables for mapping the data and for coordinating our data with other datasets. In addition to the use of ArcView software with geocoding capabilities, we will geocode using web-based companies who charge a nominal fee for processing street-level data to latitude/longitude, census block, census block group, census tract, and zip code.

For geocoding through a web-based company, a server-based software is downloaded to a desktop computer. This software connects the desktop via a secure connection to a geocoding database. A list of address is submitted via the software with no other identifying information (only addresses are submitted). A connection with the server allows for geocoding based on extensive maps held on the server. The submitted

address data does not identify the patient and is guaranteed to remain confidential based on the company's confidentiality agreement. For the confidentiality agreement of Geocode.com, please see the following website:

<http://www.na.teleatlas.com/products/geocode/ezsaleslic.html>

Visualization of the Data:

Also, as stated above, our research plan includes plotting data at aggregated levels (zip code or census tract). It is our belief that the geographic location of a health event (i.e. injury, accident, trauma) is not a unique health identifier in many cases, and therefore does not infringe on the confidentiality of the patient. Where the type of incident could become a unique identifier we will use graphic techniques to obscure the original point location of the data. First, aggregated areas such as census block, census tract, census block group, and zip code will be used to obscure the geographic location of sensitive data. Graduated colors (where each aggregated area is a different color based on the value of the assigned attribute) or randomized dot density (where each aggregated area is displayed with dots representing some value of the assigned attribute but not the actual point location) are common techniques to represent location data without revealing point location. If certain aggregate areas to be plotted contain less than 100 residents we will use the next largest area as the visual aggregate.

We will also be geocoding home addresses of all patients seen at Cooper Hospital, Our Lady of Lourdes Medical Center, and Virtua Health Camden. The data will not be presented as point locations graphically. Once geocoded the address location field will be removed from the working dataset to ensure patient privacy. The home address data will only be plotted by aggregate area (as explained above), with no area used that contains less than 100 residents.

Data Analysis Plan, Statistical Tests - The aggregate level data will be converted to incidence data using already established population levels for the neighborhoods in Camden. The primary goal of the data collection is to provide descriptive data for health indicators based on geographical location, age, sex and race/ethnicity. Once compiled and presented, data will be available for analysis by community or government for purposes of tracking health indicators over time, understanding neighborhood issues, applying for grant money, or creating policy.

Risks and Benefits

Potential Risks to Subjects and How They Will Be Minimized - There is no to minimal risk to individuals included in the study. All data obtained from the medical records will remain confidential as described above. No attempts will be made to contact individual patients.

Potential Benefits - There is significant benefit to the community for reasons already described. Ability to track health indicators over time and by neighborhood will afford the community a way of assessing changes in health.

Risk: Benefit Ratio – The benefit of tracking health indicators over time and for particular geographic areas within Camden is of great benefit to the Camden community, and to society as a whole to better understand the conditions of urban health and urban health care.

Plans for Monitoring Subject Safety

Privacy issues are the main concern with this study. No other risk is involved. Several measures to protect private patient information are included in our protocol.

Procedures to Maintain Privacy and Confidentiality

All unique identifiers will be destroyed from the data files once data is fully compiled and ready to be analyzed. Only one copy of the final database, with unique addresses, but without patient name, record, or other identifying information, will be maintained on a secure computer in the Department of Family Medicine at Cooper Hospital. Other datasets to be worked with and maintained on other computers of researchers in other departments will only contain aggregated data without unique address information. Aggregated datasets will be publicly available through CamConnect for use by community organizations, individuals, or government officials.

PHI To Be Obtained

The following chart describes the type of PHI that will be collected, which patients the information will be collected for, sources of the data, and the presentation of that data in the final public dataset. All data will be obtained by researchers listed on the study who have completed HIPAA training, the IRB-approved Human Subjects course, and are registered with the IRB. None of the following will be identified as individual datapoints, rather as aggregates of similar data.

PHI	Source (primary)	Source (secondary)	Diagnoses code limits	Form of data made public*
Pt. Name	EMR	None	All	None
Medical record #	EMR	None	All	None
Date of birth	EMR	None	All	Age (years)
Sex	EMR	None	All	Sex
Address	EMR	None	All	Neighborhood
Race	EMR	None	All	Aggregate race
Date of admission	EMR	None	All	Length of stay
Date of discharge	EMR	None	All	Length of stay
Diagnoses codes	EMR	None	All	Condition categories**
Insurance status	EMR	None	All	Medicare, Medicaid, private, uninsured
Diabetes Medications	EMR	Chart review	Diabetes codes	Diabetes medication class
HgA1C	EMR	Chart review	Diabetes codes	HgA1C
PFT	EMR	Chart review	Asthma, COPD codes	PFT
Asthma, COPD medications	EMR	Chart review	Asthma, COPD codes	Asthma, COPD medication class
Frequency of visits	EMR	Chart review	Asthma, COPD, Diabetes	Frequency of visits
ASD/MI/IHD/CVD* medications	EMR	Chart review	ASD/MI/IHD/CVD	ASD/MI/IHD/CDB medication class
Lipid panel	EMR	Chart review	ASD/MI/IHD/CVD	Lipid panel value ranges
Cardiac stress test	EMR	Chart review	ASD/MI/IHD	Cardiac stress test categorical findings
Birth weight	EMR	Chart review	Birth	Birth weight

Gestational age	EMR	Chart review	Birth	Gestational age
Prenatal/Perinatal medications	EMR	Chart review	Prenatal/perinatal	Prenatal/Perinatal medication class
Ultrasonography categorical findings	EMR	Chart review	Prenatal	Ultrasonography categorical findings
Blood lead levels	EMR	Chart review	All	Lead level ranges
Geographic location of injury	EMR	Chart review	Injuries	Geographic location on map
Assault ROS	EMR	Chart review	Injuries	Assault ROS
Month of admission	EMR	Chart review	Asthma, injuries	Month of admission
Time of day of admission	EMR	Chart review	Injuries	Time of day in blocks of 6 hours
				*All data presented as aggregate unless otherwise noted.
* ASD (atherosclerotic diseases), MI (myocardial infarction), IHD (ischemic heart disease), CVD (cerebral vascular disease)				**Condition categories will be aggregates of multiple diagnoses codes for a certain condition (i.e. asthma, COPD, diabetes, assault, STD, etc.)

These aggregate groups will be presented in tables, charts, and maps that display groups, not individuals. For instance, one of the most detailed potential tables/maps might display number of black (race) male (sex) patients, ages 0-14 (age group), who presented to the ED for falls (diagnoses code), broken down by 6 hour time blocks (time of day), but presented as a yearly figure, with the point where they fell mapped as a point location (geographic location of injury). Nowhere would this data suggest that a particular 4 year-old black male patient fell at a particular point between the hours of 6PM-12AM on a particular day. Rather, the table would show that in a certain neighborhood, a certain number of black male children between the ages of 0-14 were injured in falls during a certain year. The yearly number would also be presented broken down in the 6 hours blocks (time of day) for the group (as a whole, not individual patients), or potentially by month instead of year (if there were significance for monitoring or evaluating a program). Also, one could see on the map where these falls occurred, but that point would not be linked to specific information beyond the aggregate information for the group about age, race, sex, time of day.

PHI Protection Protocol

To ensure that all data remains secure and that all PHI remains confidential, a detailed plan for acquiring, working with, presenting, and destroying the data has been developed.

Original data files and Verified Data Files: All data acquired electronically from various electronic hospital records will be burned to individual CD-ROM discs. The data will then be transferred to a secured storage device that is locked in the department of Family Medicine at Cooper Hospital in the Education and Research Building (E&R). Once data is stored in its original form on this device, the transfer CDs will be destroyed. All data files (named as *original*, *verified*, and *working*) will be password protected and stored on the secured storage device. Only researchers who are partners of this study, who are registered with the Cooper IRB, and who have completed HIPAA training and the online Human Subjects course will have access to these files and passwords, and only on the premises of the E&R building at Cooper Hospital.

The *original data files* will contain all fields of sensitive PHI listed in the proposal and collected from the participating hospitals. The first step of work done with the *original data files* will be to create a *verified data file* that has been checked for accuracy, duplications, or other problems with the *original data file*. PHI in the *verified data files* that will remain includes all PHI that was in the *original data file*. A final copy of the *verified data file*

will be saved to the secured storage device. The *original data files* and the *verified data files* will be stored on the secured storage device for up to 2 years from the date of acquisition for purposes of research methods and quality control. At that time the data will be destroyed, and the storage device will be reformatted to ensure all data is adequately erased.

Neither of these files, the *original data files* nor the *verified data files* will leave Cooper Hospital or be saved on any disk other than the secured storage device in the E&R building at Cooper Hospital.

Working Data Files: Once all data has been verified, PHI such as patient name, medical record number, date of birth, date of admission, date of discharge, home address, and any identifiers of address below the neighborhood (census tract) level will be destroyed or de-identified by aggregation, and from that point on the *working data files* will not contain these sensitive PHI. These files will be the intermediate files that will be used to create the *final public data files*. The process of aggregating data for de-identifying is the purpose of the *working data files*. *Working data files* will be stored on the secured storage device for up to 2 years from the date of acquisition for purposes of research methods and quality control.

Final public data files: The *final data files* that will be publicly published via the department of Family Medicine, periodicals, or CamConnect, will not contain the most sensitive PHI. The PHI in these final public files that will have been destroyed or de-identified include: Patient name, medical record number, date of birth, date of admission, date of discharge, home address, and any identifiers of address below the neighborhood (census tract) level. All other PHI listed will be thus de-identified and available through aggregated groups in tables, charts, or maps that adequately protect the PHI of individuals. Data considered PHI will not be presented as individual cases, only as aggregates.

References

Billings J, G Anderson, and L Newman. *Recent Findings on Preventable Hospitalizations*. Health Affairs, 1996 Fall; 15(3).

Burr J, Sherman G, Prentice D, Hill C, Fraser V, Kollef MH. *Ambulatory care-sensitive conditions: clinical outcomes and impact on intensive care unit resource use*. South Med J. 2003 Feb; 96(2).

Forst LS. *A State Trauma Registry as a Tool for Occupational Injury Surveillance*. Journal of Occupational and Environmental Medicine, Volume 41, Number 6, June 1999.

Kellermann AL. *Community-level firearm injury surveillance: Local data for local action*. Annals of Emergency Medicine, Volume 38 № Number 4 № October 2001.

Kingsley GT. *Neighborhood Indicators: Taking advantage of the new potential*. The National Neighborhood Indicator Project, The Urban Institute. October 1998.

Kyran QP. *Expanding the National Electronic Injury Surveillance System to Monitor All Nonfatal Injuries Treated in US Hospital Emergency Departments*. Annals of Emergency Medicine, Volume 34 № Number 5 № November 1999.

Laditka JN, Laditka SB, Mastanduno MP. *Hospital utilization for ambulatory care sensitive conditions: health outcome disparities associated with race and ethnicity*. Soc Sci Med. 2003 Oct.

Parker J and Schoendorf K. *Variation in Hospital Discharges for Ambulatory Care-Sensitive Conditions Among Children*. Pediatrics Oct 2000.

Pappas G, Hadden W, Kazak LJ, Fisher G. *Potentially avoidable hospitalizations: inequalities in rates between US socioeconomic groups*. Am J Public Health. 1997;87/

Pomerantz WJ. *Relationship between socioeconomic factors and severe childhood injuries*. J Urban Health - 01-Mar-2001; 78(1): 141-51.

Shepherd JP. *Recording of violent offences by the police: an accident and emergency department perspective*. Med Sci Law. 1989;29:251-257

Wallace D. *Scales of geography, time, and population: the study of violence as a public health problem*. Am J Public Health - 01-Dec-1998; 88(12): 1853-8.