**Protocol Title:**

CONSCIENCE – CONnecting SCIence – ENgaging Chicago for Equity: A Multi‑Site Retrospective Analysis of Chronic Disease Burden Across the Chicago Metropolitan Region Using CAPriCORN Network Data

**Principal Investigator:**

**Juan C. Rojas, MD, MS**

Associate Chief Medical Information Officer

Director, Rush Health Equity Analytics Studio, RUSH BMO Institute for Health Equity

Assistant Professor, Department of Internal Medicine – Division of Pulmonary, Critical Care & Sleep Medicine, Rush University Medical Center Tel 312‑563‑2301   |   juan\_rojas@rush.edu

**Co-Investigator(s):**

**Abel N Kho, MD**

**Director, Institute for Artificial Intelligence in Medicine**

**Professor, Medicine (General Internal Medicine), Preventive Medicine (Biostatistics and Informatics)**

**Northwestern University**

**Sponsor:** Sponsor Searle Family Trust (July 1 2024 – July 1 2029)

**Version: 1.0**

**Date: June 10th, 2025**

**Engaged Study Personnel:**

1. **Jada Sherrod, MHA**

**Business Manager of Operations and Finance**

**Rush BMO Institute for Health Equity**

1. **Jason Stanghelle, MS**

**Data Analyst - Rush BMO Institute for Health Equity**

**Participating CAPriCORN Sites:**

AllianceChicago

Cook County Health

NorthShore University Health System

Northwestern University

Rush University Medical Center

University of Chicago

University of Illinois at Chicago

**Revision History**

|  |  |  |
| --- | --- | --- |
| **Revision #** | **Version Date** | **Summary of Changes** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Table of Contents

[**1.** **Project Summary/Abstract** 5](#_Toc120801810)

[**2.** **Background/Scientific Rationale** 5](#_Toc120801811)

[**3.** **Objectives/Aims** 5](#_Toc120801812)

[**3.1.** **Study Endpoints** 5](#_Toc120801813)

[**4.** **Eligibility** 5](#_Toc120801814)

[**4.1.** **Inclusion Criteria** 5](#_Toc120801815)

[**4.2.** **Exclusion Criteria** 5](#_Toc120801816)

[**4.3.** **Protected/Vulnerable Populations** 5](#_Toc120801817)

[**5.** **Number of Subjects & Statistical Considerations** 6](#_Toc120801818)

[**6.** **Study Design & Data Collection** 6](#_Toc120801819)

[**7.** **Risks and Benefits** 7](#_Toc120801820)

[**8.** **Data Management, Monitoring, Confidentiality & Analysis** 7](#_Toc120801821)

[**9.** **Informed Consent** 8](#_Toc120801822)

[**10.** **Unanticipated Problems** 8](#_Toc120801823)

[**11.** **Resources Available** 8](#_Toc120801824)

[**12.** **References** 9](#_Toc120801825)

1. Project Summary/Abstract

The Chicago metropolitan region faces a significant life expectancy gap between its most and least advantaged communities, with recent data showing disparities of up to 30 years between specific neighborhoods (Streeterville at 90 years vs. Englewood at 60 years)1. This "death gap" represents one of the starkest health inequities in America, driven primarily by the unequal distribution of chronic diseases2. Despite the severity of this crisis, Chicago lacks a comprehensive, real-time surveillance system to identify and monitor chronic disease hot zones across the entire metropolitan area.

CONSCIENCE (CONnecting SCIence -- ENgaging Chicago for Equity) will create the nation's first metropolitan-wide chronic disease dashboard with comprehensive population coverage through the CAPriCORN network. This retrospective multi-site study will analyze electronic health record data from 2019-2024 across approximately 2.8 million unique adult patients to:

1. **Build a first-of-its-kind dashboard tracking 38 chronic diseases that drive the death gap**
2. **Identify and characterize chronic disease hot zones at the census tract level**
3. **Enable direct community action through real-time data access and monitoring**

The study leverages CAPriCORN's extraordinary reach across academic medical centers, safety net hospitals, and community health centers to capture healthcare encounters from all socioeconomic strata3. Based on recent CAPriCORN data, the network contains approximately 12.8 million patient records (not deduplicated), representing about 40% of area adults with healthcare encounters. The Medical Research Analytics and Informatics Alliance (MRAIA) will serve as the honest broker, performing critical patient deduplication to ensure accurate population-level estimates. MRAIA is a non-profit organization experienced in providing data stewardship for multi-site studies and public health projects4.

The resulting dashboard will transform public health practice by providing neighborhood-level chronic disease intelligence updated every six months. Community organizations in identified hot zones will receive tailored data packets and support for intervention design. This innovative surveillance system will enable precision public health interventions, track progress in near real-time, and serve as a model for other metropolitan areas facing similar health disparities.

1. Background/Scientific Rationale

**The Death Gap Crisis in Chicagoland**

Recent analyses confirm a life expectancy gap in Chicago ranging from 20-30 years between neighborhoods5. The most widely cited figure comes from a 2019 study by the NYU School of Medicine, showing a 30-year gap between Streeterville (90 years) and Englewood (60 years). The Chicago Department of Public Health's 2024 data show the Black-white life expectancy gap at 11.4 years overall, with some individual neighborhoods showing even larger disparities6.

Chicago also faces a public health crisis of firearm violence that disproportionately affects the same communities burdened by chronic disease. In 2023, Chicago recorded over 2,800 shooting victims, with young Black men aged 18-34 representing over 60% of victims despite being less than 5% of the population7. However, focusing only on shootings vastly underestimates the true health burden of firearm violence. For every firearm death, there are approximately 2-3 non-fatal injuries requiring emergency care, plus countless follow-up visits, surgeries, rehabilitation services, and long-term care for permanent disabilities. The psychological trauma extends even further, affecting entire communities.

While individual healthcare systems track their own patients, no comprehensive surveillance system exists to capture the full health impact of firearm violence, from initial trauma through long-term sequelae, nor one that examines how violence exposure interacts with and exacerbates chronic disease burden in affected communities.

**The Missing Tool: A First-of-Its-Kind Regional Dashboard**

Despite the severity of the death gap, public health officials and community organizations lack a critical tool: a comprehensive, neighborhood-level chronic disease dashboard that captures the vast majority of healthcare encounters across all of Chicagoland.

Current surveillance methods are fragmented and outdated:

* Death certificates provide data years after the fact.
* Hospital discharge data miss outpatient care, where most chronic disease management occurs.
* Surveys capture only small samples, missing neighborhood-level patterns.
* Individual health system dashboards create blind spots regarding community-level burden.

This fragmentation prevents effective resource allocation and targeted intervention design.

**The Power of CAPriCORN's Comprehensive Coverage**

The CAPriCORN (Chicago Area Patient-Centered Outcomes Research Network)3currently includes data on approximately 12.8 million patient records across its network partners. As a PCORnet Clinical Data Research Network, CAPriCORN uses the standardized PCORnet Common Data Model8, enabling:

* Coverage of diverse patient populations across all socioeconomic strata
* Capture of healthcare encounters from academic centers to community clinics
* Established data infrastructure with proven deduplication capabilities
* Track record of multi-site collaboration since 2013

This study builds on validated chronic disease algorithms (e.g., Elixhauser9 codes) using ICD-10-CM codes, consistent with PCORnet Common Data Model specifications and validated administrative data methodologies for chronic disease surveillance.

1. Objectives/Aims

**Aim 1. Build and Deploy a Comprehensive Chronic Disease Surveillance Dashboard for Chicagoland**

We will develop and operationalize a data platform integrating healthcare encounters across ~2,100 census tracts, covering 38 chronic diseases linked to the region’s death gap.

* Hypothesis: Integrating multisite patient data and applying advanced spatial analyses will identify statistically significant disease clusters and enable more accurate estimation of chronic disease burden.
* Impact: This aim will produce the first near-real-time regional surveillance tool for chronic disease disparities in a major U.S. city, setting a national precedent.

Sub-aim 1.1. Quantify the Impact of Multi-Site Deduplication on Disease Prevalence Estimates

Apply MRAIA-based deduplication to patient records across health systems to refine prevalence calculations.

* Hypothesis: 15–25% of patients receive care across multiple systems, with >30% among those with complex conditions, leading to inflated prevalence without deduplication.
* Impact: This sub-aim will improve the methodological rigor of metropolitan disease surveillance and produce a replicable approach for multi-institutional health data integration.

Sub-aim 1.2. Analyze Comprehensive Firearm Violence Health Burden as a Contributor to the Death Gap

Use ICD-10-CM codes to track all firearm-related encounters—acute, follow-up, and long-term complications.

* Hypothesis: The health burden of firearm violence is significantly underestimated when mortality is the sole metric; comprehensive tracking will reveal long-term impacts that align geographically with chronic disease hot zones.
* Impact: Reframing firearm violence as a chronic public health condition will expand the scope of intervention strategies and allow integration with chronic disease prevention efforts.

**Aim 2. Identify and Characterize Chronic Disease Hot Zones**

We will apply spatial clustering techniques to define “hot zones” of chronic disease burden and characterize them across clinical, demographic, and geographic dimensions.

* Hypothesis: Chronic disease hot zones are spatially stable and defined by recurring patterns of multimorbidity, demographic vulnerability, and limited access to care.
* Impact: By profiling disease clusters in detail, this aim will inform localized intervention strategies tailored to community-specific needs and conditions.

**Aim 3. Translate Hot Zone Data into Community Action to Reduce the Death Gap**

We will engage community-based organizations in hotspot regions to co-design interventions informed by data and evaluate their health impact.

* Hypothesis: Providing community partners with timely, geographically specific data and supporting tailored interventions will reduce the chronic disease burden and life expectancy gap in target areas.
* Impact: This aim will demonstrate a replicable “Data-to-Action” model linking high-resolution analytics to equitable public health improvement.

## 3.1. Study Endpoints

**Primary Endpoints:**

* Age-standardized chronic disease prevalence rates by census tract for 38 validated conditions
* Identification of statistically significant disease hot zones using spatial scan statistics (p<0.05)
* Quantification of patient overlap across health systems and its impact on prevalence estimates (% deduplicated)
* Comprehensive firearm violence health burden, including:
  + Incidence rates of initial firearm encounters
  + Prevalence of patients with firearm-related sequelae
  + Total healthcare utilization for firearm-related care (initial + subsequent encounters)

**Secondary Endpoints:**

* Temporal trends in chronic disease burden (annual percent change 2019-2024)
* Within-tract demographic disparities in disease prevalence (rate ratios by race/ethnicity)
* Geographic clustering patterns of multi-morbidity (≥3 chronic conditions)
* Community engagement metrics (# organizations engaged, interventions launched, residents reached)

1. Eligibility

### 4.1. Inclusion Criteria

* Age ≥ 18 years at time of any healthcare encounter
* At least one ambulatory, emergency department, or inpatient visit at a participating CAPriCORN site during the study period (January 1, 2019 – December 31, 2024)
* Valid geocoded address within the Illinois portion of the Chicago metropolitan region, including:
  + Cook County (all 1,319 census tracts)
  + DuPage County (all 219 census tracts)
  + Lake County, IL (all 176 census tracts)
  + Will County (all 168 census tracts)
  + Kane County (all 121 census tracts)
  + McHenry County (all 77 census tracts)

### 4.2. Exclusion Criteria

* Age < 18 years at all encounters
* No valid geocoded address available (expected <5% based on CAPriCORN geocoding rates)
* Addresses outside Illinois
* Addresses outside the six-county Chicago metropolitan region
* Encounters occurring outside the study period

### 4.3. Protected/Vulnerable Populations

**Individuals who are not yet adults (infants, children, teenagers):**

**Excluded** - study limited to adults ≥18 years to focus on chronic disease burden in adult populations

**Pregnant women:** **Included incidentally** - pregnancy status will not be identified in the aggregate data. Pregnancy-related conditions are not among the 38 chronic diseases tracked.

**Prisoners:** **Included incidentally** - incarceration status will not be identified in the aggregate data. Cook County Jail healthcare encounters through Cermak Health Services (part of Cook County Health) may be included.

While these populations may be included in the healthcare encounters captured, their special status will not be identified or analyzed separately as all data is aggregated at the census tract level with minimum cell size of 10.

**Network Coverage Ensures Representative Surveillance:**

1. **Academic Medical Centers** (5 institutions)
   * Quaternary care for complex chronic diseases
   * Specialty clinics for rare conditions
   * Clinical trial participants
2. **Safety Net Systems** (3 institutions)
   * Cook County Health: Largest public health system in Illinois
   * AllianceChicago: Network of 70+ FQHCs
   * UI Health: Safety net academic center
3. **Community Health Systems** (2 institutions)
   * Endeavor Health: Suburban community hospitals
   * Multiple affiliated community practices
4. **Data Infrastructure**
   * MRAIA: experience as honest broker
   * Proven deduplication algorithms
   * HIPAA-compliant data processing

**Population Coverage**: This collaboration ensures capture of:

* All socioeconomic strata (from Medicaid to commercial insurance)
* All major racial/ethnic communities in proportions reflecting Chicago demographics
* Urban core and suburban populations
* Care settings from FQHCs to quaternary care centers

1. Number of Subjects & Statistical Considerations

### 5.1 Expected Sample Size

CAPriCORN's nine-system data warehouse contains approximately **12.8 million patient records** (as of 2021, undeduplicated). After cross-site deduplication by MRAIA, adult restriction (≥18 years), and limiting to the 2019-2024 study window, we estimate **approximately 2.8 million unique adults** eligible for analysis.

**Population Coverage Context**: U.S. Census data show roughly **7.0 million adults** among the **8.98 million total Chicago-metro residents in 2024**; thus approximately 40% of area adults have CAPriCORN encounters during the study period, providing robust population-level estimates.

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject Category** | **Estimated N** | **Data Source** | **Notes** |
| Adults (≥ 18 y) | ~2.8 M | CAPriCORN post-deduplication | Primary analytic cohort |
| Children (< 18 y) | Excluded | — | Outside study scope |
| Pregnant Women | Incidental inclusion | CAPriCORN | Status not identified in aggregate data |
| Prisoners | Incidental inclusion | CAPriCORN | Status not captured; may include Cook County Jail |

### 5.2 Primary Outcome & Effect Size

The **primary endpoint** is the **age-standardized prevalence** of 38 chronic diseases at the **census-tract-year level**, calculated using direct standardization to the **2020 U.S. Census population**10. **Hot zones** are identified using **Local Moran’s I**,11 with **FDR-adjusted p < 0.05**, signifying statistically significant spatial clustering. **Secondary endpoints** include spatial patterns of **multimorbidity**, the full **health system burden of firearm violence**, and improved prevalence estimates via **multi-system deduplication**. **Intervention effects** will be assessed through **difference-in-differences analysis**, comparing changes in disease burden between intervention and matched control tracts. With **twice-yearly data updates**, these endpoints enable near real-time tracking, cluster detection, and timely evaluation of localized public health interventions.

### 5.3 Power Analysis

* Prevalence difference detection: Assuming a mean tract population of 3,000 adults and baseline prevalence of 10%, we have >99% power (α = 0.05, two-tailed) to detect a 2-percentage-point absolute difference between equally sized tract groups (1,050 tracts each).
* Spatial cluster detection: Monte Carlo simulations12 (10,000 iterations) inserting synthetic clusters show >95% sensitivity to detect hot zones of ≥900 adults (~0.03% of the full cohort), ensuring sensitivity to even small-area concentrations.
* Temporal trend detection: With 6 annual time points per tract, we have >95% power to detect a 5% annual change in prevalence using joinpoint regression

### 5.4 Precision of Estimates

The resulting dataset comprises approximately 12,600 tract-year cells (2,100 census tracts × 6 years). After applying statistical disclosure control—suppressing cells with fewer than 10 observations—we anticipate that over 98% of cells will remain analyzable, yielding robust and precise estimates:

* 95% confidence intervals of ±0.3 to 0.6 percentage points for conditions with 5–10% prevalence
* Coefficient of variation (CV) <10% for highly prevalent conditions such as diabetes and hypertension
* Stable subgroup estimates (by age, sex, or race/ethnicity) in more than 90% of census tracts

### 5.5 Geographic Distribution

**Expected sample distribution across the six-county region:**

|  |  |  |  |
| --- | --- | --- | --- |
| **County** | **Population** | **Census Tracts** | **Expected Patients** |
| Cook | ~5.2M | 1,319 | ~1.8M |
| DuPage | ~930K | 219 | ~320K |
| Lake | ~700K | 176 | ~240K |
| Will | ~690K | 168 | ~240K |
| Kane | ~530K | 121 | ~180K |
| McHenry | ~310K | 77 | ~110K |
| **Total** | **~8.4M** | **2,080** | **~2.8M** |

1. Study Design & Data Collection

**Study Design:**

Retrospective, multi-site, ecological study analyzing aggregate population-level data to create a chronic disease surveillance system.

**Data Flow Structure:**

* **Site to Honest Broker**: CAPriCORN sites submit a limited dataset containing patient-level data to MRAIA (honest broker) for cross-site deduplication
* **Honest Broker to PI**: MRAIA processes the limited dataset to remove duplicate patients across sites, then provides the PI/research team with **only aggregate data** - summary counts by census tract, year, and demographics
* **PI Access**: The PI and research team never receive or access the limited dataset; all analyses are performed exclusively on aggregate population data

This two-stage ecological design enables accurate population-level surveillance by preventing duplicate counting across healthcare systems while ensuring the research team works only with fully de-identified aggregate data. The design is appropriate for identifying geographic patterns of disease burden and high-risk areas ("hot zones") without any exposure to individual patient information.

**Data Collection Overview:**

**Type of Data:**

* **Temporality**: Retrospective - all data exists at time of protocol submission (2019-2024)
* **Level**: Individual-level data (raw extraction) will be accessed by local Data Analysts with authorization to access institutional data warehouses. This will be used as input for MRAIA. The resulting datasets used for analysis will consist ofaggregate data only - no individual-level data received by investigators after transformation.

**Raw Data Variables Requested and resulting dataset format (Aggregate Level):**

|  |  |  |  |
| --- | --- | --- | --- |
| **Domain** | **Raw data element (individual-level)** | **Transformed data element (aggregate-level)** | **Rationale for inclusion** |
| **Disease counts** | ICD-10-CM Codes | Number of patients with each of 38 chronic conditions (see Appendix A) | Necessary for disease surveillance (Aim 1) and identification of chronic disease hot zones (Aim 2) |
| Total number of encounters (Denominators) | Total unique patients for prevalence/rate calculations | Critical for prevalence/rate calculations |
| **Firearm violence counts** | ICD-10-CM Codes | Number of patients with ANY firearm-related encounter (initial, subsequent, or sequela)  See Appendix B for firearm-related violence ICD-10-CM codes. | Inclusion of firearm violence data will allow the study’s unique exploration of Firearm violence as a primary driver of the racial life expectancy gap |
| **Demographics** | Age\*  \*ages 90 and above aggregated into single category in the patient-level extract performed locally at each site | Counts for age groups: 18-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ | Required for age-standardized chronic disease prevalence rates. |
| Sex | Counts for sex categories: Male, Female, Unknown |  |
| Race | Counts for race categories: White, Black/African American, Asian, Native American/Alaska Native, Native Hawaiian/Pacific Islander, Multiple Races, Other, Unknown | Required for measuring demographic disparities in disease prevalence, one of the study’s secondary endpoints. |
| Ethnicity | Counts for ethnicity categories: Hispanic/Latino, Non-Hispanic/Latino, Unknown | Required for measuring demographic disparities in disease prevalence, one of the study’s secondary endpoints. |
| **Geographic and Temporal Variables** | Census tract identifier (FIPS-11 code) | Counts for FIPS-11 codes | In addition to enabling regional surveillance for communities, these are required for identification of characterization of chronic disease hot zones (Aim 2). |
| Year (2019, 2020, 2021, 2022, 2023, 2024) | Counts for each year | Required for measuring Temporal trends in chronic disease burden (annual percent change 2019-2024), one of the study’s secondary endpoints. |

**HIPAA Identifiers:**

☐ Names

X Address (including street address, city, county, precinct, or 5-digit zip code)

☐ Any elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death, date of office visit

☐ Telephone numbers

☐ Fax numbers

☐ Electronic mail addresses

☐ Social security numbers

☐ Medical record numbers

☐ Health plan beneficiary numbers

☐ Account numbers

☐ Certificate/license numbers

☐ Vehicle identifiers and serial numbers, including license plate numbers

☐ Device identifiers and serial numbers

☐ Web Universal Resource Locators (URLs)

☐ Internet Protocol (IP) address numbers

☐ Biometric identifiers, including finger and voice prints

☐ Full face photographic images and any comparable images

☐ OTHER unique identifying number, characteristic, or code: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**NOTE:** The HIPAA identifier indicated above (geospatial data) will be present in the Patient-Level Extract that is generated locally at each site and provided to MRAIA. The final aggregate dataset returned to the investigators from MRAIA will not include any HIPAA identifiers. All geographic identifiers are limited to census tract level (typically 1,200-8,000 residents) and will be rolled up into counts returned to the investigator Census tracts at the patient-level will not be returned to the investigator.

**Study Period:**

Medical information from January 1, 2019, to December 31, 2024 (6 complete years) is initially pulled, but the refresh continues until July 1, 2029

**Query Timing:**

* **Initial query**: One-time extraction covering full study period (Summer 2025)
* **Refresh queries**: Every 6 months through July 1, 2029 for dashboard updates
* **Justification**: Semi-annual updates balance timeliness with operational feasibility

**Interventions:**

None - purely observational study

**Honest Data Broker:**

Medical Research Analytics and Informatics Alliance (MRAIA) - serving as honest broker since CAPriCORN's inception with established BAAs with all sites

**Detailed Data Flow Process:**

**Data Architecture Overview:**

STEP 1: Data Standardization at Each Site

Site EHRs → PCORnet CDM

- Each CAPriCORN site converts their electronic health records into the standardized PCORnet Common Data Model format

STEP 2: Data Extraction

PCORnet CDM → Limited Dataset Extraction

- Sites query their PCORnet database to extract patient-level data with minimal identifiers needed for deduplication

STEP 3: Honest Broker Processing

Limited Dataset → MRAIA (Honest Broker)

- MRAIA receives limited datasets from all sites

- Performs cross-site patient deduplication (removes patients who appear at multiple hospitals)

- Generates aggregate summary counts by census tract, year, and demographics using patient data from the last visit location during the study period for that year or 6 month window for subsequent data pulls

- Destroys all patient-level data after aggregation

STEP 4: Research Team Access

MRAIA → Aggregate Data Only → Rush Analytics Studio

- PI/research team receives ONLY the aggregate counts

- No patient-level data ever reaches the research team

- Data is loaded into secure Azure Analytics environment at Rush

STEP 5: Public Dissemination

Rush Analytics Studio → Azure Analytics Platform → Public Dashboard

- Aggregate data is analyzed and visualized

- Results are published on publicly accessible dashboard

- All data remains at population level with privacy protections

**PHASE 1: Site-Level Extraction (Weeks 1-4)**

Each participating site will:

1. **Execute Standardized Query Package**
   * SQL code provided by study team validated against PCORnet CDM v7.0
   * Extracts from standardized tables: DIAGNOSIS, DEMOGRAPHIC, ENCOUNTER
   * Creates two output files per site
2. **Generate Patient-Level Extract (COMORBIDITIES\_DEMOGRAPHICS\_FIPS11.csv)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CAP\_ID** | **YEAR** | **FIPS11** | **AGE** | **SEX** | **RACE** | **HISPANIC** | **HYPERTENSION** | **CHD** | **CHF** | **... [38 conditions]** |
| 12345 | 2019 | 17031123456 | 67 | M | BLACK | N | 1 | 0 | 1 | ... |
| 12345 | 2020 | 17031123456 | 68 | M | BLACK | N | 1 | 1 | 1 | ... |

CAP\_ID is a study-specific identifier created at each site, not linked to medical record numbers.

**PHASE 2: Secure Transfer to MRAIA (Week 5)**

* Transfer via MRAIA's secure FTP with multi-factor authentication
* MRAIA confirms receipt

**PHASE 3: MRAIA Processing (Weeks 6-7)**

MRAIA performs the following validated steps:

1. **Patient Matching and Deduplication**
   * Creates unified patient identifier across all sites
2. **Geographic Validation**
   * Confirms FIPS codes correspond to valid Illinois census tracts
   * Excludes out-of-state addresses using geographic boundaries
3. **Hierarchical Aggregation Process**

1. By Census Tract (2,080 tracts in Chicago metropolitan area)

↓

2. By Year (6 years: 2019, 2020, 2021, 2022, 2023, 2024)

↓

3. By Condition (38 chronic diseases + firearm violence categories)

↓

4. By Demographics:

• Age Groups (7 categories): 18-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+

• Sex (3 categories): Male, Female, Unknown

• Race (8 categories): White, Black/African American, Asian, Native American/Alaska Native,

Native Hawaiian/Pacific Islander, Multiple Races, Other, Unknown

• Ethnicity (3 categories): Hispanic/Latino, Non-Hispanic/Latino, Unknown

↓

5. Apply Privacy Protection: Cell suppression for any count <10

**PHASE 4: Aggregate Data Return (Week 8)**

MRAIA returns comprehensive stratified aggregate data files:

**Primary Output: GEO\_AGGREGATES\_STRATIFIED.csv**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **FIPS11** | **YEAR** | **CONDITION** | **AGE\_GROUP** | **SEX** | **RACE** | **ETHNICITY** | **COUNT** |
| 17031010100 | 2019 | Diabetes | 45-54 | F | BLACK | NON\_HISPANIC | 127 |
| 17031010100 | 2019 | Diabetes | 45-54 | F | BLACK | HISPANIC | 43 |
| 17031010100 | 2019 | Diabetes | 55-64 | M | WHITE | NON\_HISPANIC | 89 |

**PHASE 5: Semi-Annual Refresh (through July 1, 2029)**

* Sites run updated extracts every 6 months
* MRAIA processes incremental updates
* Dashboard automatically refreshes with new data
* Community partners notified of updates

All data remains aggregate throughout the process. Investigators never access individual-level data, ensuring privacy protection exceeding HIPAA requirements.

1. Risks and Benefits

**Risk Level:**

No more than minimal risk - consistent with 45 CFR 46.104(d)(4) for secondary research with identifiers removed

**Detailed Risk Assessment:**

**1. Privacy Risk from Geographic Identification**

* **Risk**: Potential identification of individuals in sparsely populated census tracts
* **Likelihood**: Very low (<0.01%)
* **Mitigation strategies**:
  + Census tracts contain 1,200-8,000 residents (median ~3,000)
  + Cell suppression for any stratified count <10
  + Complementary suppression to prevent back-calculation
  + Secondary suppression of demographically unique cells
* **Context**: Census tract data routinely published by CDC, Census Bureau for health surveillance

**2. Risk from Condition Disclosure**

* **Risk**: Inference about individual health conditions in small geographic areas
* **Likelihood**: Negligible
* **Mitigation strategies**:
  + Only aggregate counts reported, never individual diagnoses
  + 38 conditions tracked simultaneously, preventing targeted inference
  + Multiple demographic strata further dilute identification risk
  + No rare diseases included that could identify individuals

**3. Risk from Multi-Site Data Linkage**

* **Risk**: Concerns about comprehensive health profiles from multiple systems
* **Likelihood**: Not applicable
* **Mitigation strategies**:
  + MRAIA performs all linkage in secure environment
  + Investigators receive only post-aggregation data
  + No ability to re-link to individuals

**4. Risk of Stigmatization of Communities**

* **Risk**: Hot zone identification could stigmatize neighborhoods
* **Likelihood**: Low to moderate
* **Mitigation strategies**:
  + Community engagement in interpretation and messaging
  + Focus on structural factors, not individual behaviors
  + Paired with resource allocation, not punitive measures
  + Success stories highlighted to counter stigma
  + Firearm data presented as public health issue, not crime statistics

**5. Risk from Firearm Injury Data**

* **Risk**: Sensitive nature of violence-related health data
* **Likelihood**: Low with proper safeguards
* **Mitigation strategies**:
  + Enhanced suppression threshold (n<5) for firearm data
  + No individual-level details or incident specifics
  + Aggregate reporting only at the annual level to do less censoring
  + Focus on health outcomes

**Benefits Assessment:**

**Individual Benefits:**

None directly - retrospective aggregate analysis

**Societal and Public Health Benefits:**

**1. Comprehensive Disease Surveillance Tool**

* Creates a unique one of a kind metropolitan chronic disease dashboard with ~40% population coverage
* Provides actionable intelligence updated every 6 months
* Transforms reactive healthcare to proactive community health
* **Significance**: Addresses critical gap identified in Healthy Chicago 2025

**2. Direct Death Gap Reduction Through Hot Zone Targeting**

* Identifies specific census tracts with highest disease burden
* Enables precision allocation of limited public health resources
* Creates accountability through transparent progress tracking

**3. Community Empowerment with Actionable Data**

* Communities receive neighborhood-specific chronic disease profiles
* Data supports grant applications and advocacy efforts
* Success stories shared across communities accelerate progress
* **Innovation**: Shifts power dynamic from institutions to communities

**4. Methodological Advancement for Health Equity**

* Demonstrates feasibility of metropolitan-scale EHR surveillance
* Creates replicable framework for other cities (NYC, LA, Houston)
* Establishes new standard for public health infrastructure

**Risk-Benefit Assessment**: The minimal privacy risks (mitigated through multiple safeguards) are far outweighed by the substantial public health benefits. The study addresses Chicago's most pressing health equity challenge with innovative methods while maintaining privacy protections exceeding current standards. IRB approval is warranted given the favorable risk-benefit ratio and alignment with public health priorities.

1. Data Management, Monitoring, Confidentiality & Analysis

**Risk Level:**

Minimal risk - secondary analysis of de-identified aggregate data

**Data Governance Structure:**

**Data Governance Committee**

* **Chair**: PI (Dr. Rojas)
* **Members**:
  + Cap PI (Dr. Kho)
  + Biostatistician
  + Community representative
  + CAPriCORN Data Manager (Noah)
* **Meetings**: Quarterly
* **Responsibilities**:
  + Oversee data quality metrics
  + Review access logs quarterly
  + Approve any protocol modifications
  + Address any privacy concerns

**Technical Infrastructure:**

**Secure Analytical Environment**

* **Primary**: Rush Health Equity Analytics Studio
  + HIPAA-compliant on-premises servers
  + Isolated network segment with a firewall
  + Automated intrusion detection provided by Rush cyber team
* **Cloud**: Microsoft Azure
  + HIPAA BAA in place
  + Encrypted storage and compute

**Access Controls**

* Multi-factor authentication required
* Role-based permissions (see Access Control Matrix)
* Session timeout after 15 minutes
* All access logged with automated anomaly detection

**Access Control Matrix**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Role** | **Raw Aggregate Data** | **Analytics Platform** | **Internal Dashboard** | **Public Dashboard** |
| PI | Full access | Full access | Full access | Full access |
| Analysts | Read only | Full access | Full access | View only |
| CAPriCORN Partners | None | None | View with login | View only |
| Community Partners | None | None | Neighborhood data only | View only |
| Public | None | None | None | View only |

**Data Security Measures:**

**Technical Safeguards**

* **Encryption**: AES-256 at rest, TLS 1.3 in transit
* **Key Management**: Azure Key Vault with HSM
* **Backup**: Daily encrypted backups to a separate location
* **Disaster Recovery**: RPO 24 hours, RTO 4 hours

**Administrative Safeguards**

* Annual HIPAA training for all personnel
* Signed confidentiality agreements
* Incident response plan with 24-hour notification

**Physical Safeguards**

* 24/7 security monitoring of on-premise server rooms
* Environmental controls (temperature, humidity)
* Uninterruptible power supply

**Data Quality Assurance Framework:**

**Automated Quality Checks**

1. **Completeness**: Flag tracts with >10% missing data
2. **Consistency**: Year-over-year changes >50% trigger review
3. **Validity**: Prevalence rates compared to national estimates
4. **Accuracy**: Cross-validation with CDPH vital statistics
5. **Firearm data validation**: Cross-check with the Illinois Department of Public Health's public data portal (aggregate only).

**Statistical Disclosure Control**

* Standard suppression: n<10 for chronic diseases
* Enhanced suppression: n<5 for firearm injuries
* No reporting of exact counts between 1-9 for any conditions

**Statistical Analysis Plan:**

**Software and Computing Environment**

* **Primary**: R 4.3+ with reproducible environment (renv)
* **Packages**: sf, spdep, surveillance, epitools, tidyverse
* **Version Control**: Git with protected main branch
* **Computing**: 32-core server with 256GB RAM for spatial analyses

**Primary Analyses**

**1. Prevalence Estimation**

**For each census tract, year, and chronic condition, we will:**

1. **Calculate the crude prevalence rate** by dividing the total number of cases by the total population in that tract
2. **Compute age-standardized rates** using direct standardization, which adjusts for age differences between census tracts by applying each tract's age-specific rates to a standard population (Chicago's 2020 age distribution). This allows fair comparison between areas with different age compositions - for example, comparing a tract with many elderly residents to one with younger populations
3. **Generate confidence intervals** using the Wilson score method, which provides more accurate intervals for rates based on small numbers. This produces a lower and upper bound for each prevalence estimate, indicating the statistical uncertainty in our calculations

**Result:** Each census tract will have both crude and age-adjusted prevalence rates with confidence intervals for all 38 chronic conditions, enabling valid geographic comparisons that account for demographic differences between neighborhoods. Age standardization is crucial because chronic disease prevalence varies substantially by age, and Chicago's neighborhoods have very different age compositions.

**2. Spatial Hot Zone Detection**

We will identify geographic clusters of high disease burden using multiple validated spatial statistical methods:

* Local Moran's I statistic will detect spatial autocorrelation by identifying census tracts with high prevalence that are surrounded by other high-prevalence tracts, revealing neighborhood clusters of disease
* *Getis-Ord Gi analysis*\* will pinpoint statistically significant hot spots (areas of high prevalence) and cold spots (areas of low prevalence) by comparing each tract's rate to its surrounding neighborhood
* Kulldorff's spatial scan statistic will confirm the location and statistical significance of disease clusters using a moving circular window that tests all possible geographic cluster locations and sizes
* False Discovery Rate (FDR) correction will be applied to control for multiple testing, ensuring that no more than 5% of identified clusters are false positives

**3. Temporal Trend Analysis**

We will examine changes in disease patterns over the 6-year study period through:

* Joinpoint regression models that identify specific time points where disease or firearm injury trends significantly change direction or rate of change, revealing the impact of events or interventions
* Annual Percent Change (APC) calculations that quantify the year-over-year rate of increase or decrease for each chronic condition and firearm injury category
* Interrupted time series analysis to evaluate whether violence prevention programs or other interventions coincide with changes in firearm injury rates or chronic disease patterns in targeted areas
* Seasonal decomposition to separate underlying trends from seasonal variations in both healthcare utilization and violence patterns, accounting for predictable monthly fluctuations

**4. Disparity Analysis**

We will quantify health inequities across demographic groups and geographic areas by calculating:

* Rate ratios comparing disease and injury prevalence between demographic groups (e.g., Black vs. White residents, males vs. females) for all 38 conditions and firearm violence
* Concentration indices that measure the degree of inequality in disease burden across socioeconomic gradients, quantifying whether conditions disproportionately affect disadvantaged populations
* Intersectionality analyses examining how multiple demographic factors combine (e.g., young Black males in specific neighborhoods) to create compounded disparities in health outcomes
* Blinder-Oaxaca decomposition to identify which factors (age, location, healthcare access) explain the observed disparities between groups, distinguishing avoidable inequities from demographic differences

**Sensitivity Analyses**

We will test the robustness of our findings by examining:

1. Cell suppression impact - comparing results with and without suppressed cells to ensure privacy protections don't bias findings
2. Deduplication threshold effects - testing different patient matching criteria to verify MRAIA's deduplication accuracy
3. Geocoding precision influence - assessing how address matching quality affects geographic pattern detection
4. Geographic scale stability - confirming findings remain consistent when analyzed at different spatial units (ZIP codes, community areas)

**Data Retention and Destruction**

All study materials will be managed according to institutional policies:

* Aggregate data files will be retained for 7 years after study completion per Rush research data policy, then securely destroyed
* Analysis code and scripts will be permanently maintained in a public GitHub repository to ensure reproducibility
* Study documentation and protocols will be permanently archived in institutional repositories
* Public dashboard and visualizations will remain permanently accessible as a community resource with semi-annual updates through 2029

**Breach Response Protocol**

In the unlikely event of a data security incident:

1. Immediate response includes containment of the breach, securing affected systems, and preliminary assessment of scope
2. CHAIRb notification within 24 hours of discovery, including initial assessment and immediate actions taken
3. Site notification to all affected CAPriCORN sites within 72 hours with details of the incident and potential impact
4. Root cause analysis conducted to identify security gaps and implement corrective measures to prevent recurrence
5. Compliance documentation maintained for regulatory requirements and institutional review, with lessons learned incorporated into security protocols

While the aggregate nature of data minimizes breach risk, comprehensive security measures ensure protection exceeding HIPAA requirements.

1. Informed Consent

**Waiver of Informed Consent Requested per 45 CFR 46.116(f)**

We request a waiver of informed consent based on the following regulatory criteria:

**1. The research involves no more than minimal risk to subjects**

* **No patient contact**: Purely retrospective analysis of existing data
* **Aggregate data only**: Individual-level data never accessed by investigators
* **Multiple privacy protections**:
  + Geographic precision limited to census tract (~3,000 residents)
  + Statistical disclosure control (cell suppression)
  + No rare conditions that could identify individuals
* **Precedent**: Similar waivers routinely granted for public health surveillance

**2. The waiver will not adversely affect rights and welfare**

* **No impact on clinical care**: Retrospective analysis only
* **No additional data collection**: Uses only routine clinical data
* **Community benefit**: Results will improve public health planning in affected communities
* **Transparency**: Aggregate results publicly available on dashboard
* **Opt-out preserved**: CAPriCORN maintains opt-out registry for patients who decline research use

**3. The research could not practicably be carried out without a waiver**

**Scale Makes Individual Consent Impossible**:

* **Population**: ~2.8 million unique adults across 6 counties
* **Geographic dispersion**: 2,080 census tracts across 8,400 square miles

**Bias Introduction**:

* Consent rates vary by demographics and health status
* Would systematically exclude vulnerable populations
* Defeats purpose of comprehensive surveillance

**4. Subjects will not be provided with additional information**

* Pure retrospective analysis with no recontact
* No individual results to return
* The public dashboard provides community-level information

**HIPAA Waiver Requested per 45 CFR 164.512(i)**

We request a waiver of HIPAA authorization based on:

**1. Use involves no more than minimal risk to privacy**

**Quantitative Risk Assessment**:

* **Individual identification risk**: <0.001% based on:
  + Minimum cell size (n=10) in 3,000-person tract
  + 38 conditions tracked (not unique identifiers)
  + Multiple demographic strata
* **No PHI in final dataset**: Only aggregate counts
* **Established precedent**: CDC WONDER, County Health Rankings use similar methods

**2. Adequate protections for health information**

**Technical Protections**:

* MRAIA as experienced honest broker (15+ years)
* Hash tokens for matching (one-way encryption)
* Destruction of crosswalk after deduplication
* No retention of individual-level data

**Administrative Protections**:

* BAAs with all participating sites
* Data Use Agreements for aggregate data
* Regular audits of access and use
* Prohibited uses clearly defined

**3. Research could not practicably be conducted without PHI**

**PHI Needed for Accurate Surveillance**:

* **Geographic identifiers**: Required for census tract assignment
* **Year the encounter took place:** Needed for temporal trends. Only the year will be included. Exact dates of service (Month, Day) will NOT be used.
* **Demographics**: Essential for disparity analysis
* **Diagnosis codes**: Core to chronic disease identification

PHI used only for initial processing by sites/MRAIA. Investigators receive only de-identified aggregate data.

These waivers align with public health surveillance activities while providing stronger protections through the research framework. The aggregate nature of the final data provides additional privacy protection beyond typical surveillance systems.

1. Unanticipated Problems

Unanticipated problems involving risks to subjects or others will be reported to CHAIRb according to the following procedures:

**Reporting Timeline:**

* **Immediate** (within 24 hours): Any data breach or unauthorized access
* **Prompt** (within 5 business days): Other unanticipated problems
* **Routine** (at continuing review): Minor protocol deviations

**Examples of Reportable Problems:**

1. **Data Security Events**
   * Unauthorized access to aggregate data files
   * Malware or ransomware affecting study systems
   * Lost or stolen devices containing study data
2. **Privacy Concerns**
   * Community concerns about stigmatization
   * Attempts to re-identify individuals from aggregate data
3. **Technical Failures**
   * Geocoding errors systematically biasing certain areas
4. **Community Impact**
   * Discrimination based on hot zone identification
   * Negative community reaction requiring intervention
   * Misuse of firearm injury data by media or other entities
   * Unintended stigmatization of neighborhoods with high violence rates

**Response Protocol:**

1. **Immediate Actions**
   * Contain the problem (e.g., disable access, take the system offline)
   * Assess scope and impact
   * Notify PI and institutional officials
2. **Formal Report to CHAIRb** will include:
   * Description of the problem
   * When and how it was discovered
   * Scope (number of records, communities affected)
   * Root cause analysis
   * Corrective actions taken
   * Preventive measures implemented
   * Assessment of impact on study validity
3. **Communication Plan**
   * CHAIRb notification per institutional requirements
   * Community partner notifications if applicable, especially violence prevention organizations
   * Public disclosure if dashboard affected
   * Special protocol for any firearm data concerns with immediate community partner consultation
4. Resources Available

**Institutional Infrastructure:**

**Rush Health Equity Analytics Studio**

* Dedicated 1000sq ft secure facility
* High-performance computing cluster (512 cores, 4TB RAM) provided by Rush IT in secure server room
* Established data pipelines for health equity research
* Prior experience with census tract-level analyses

**CAPriCORN Network Infrastructure**

* Established data sharing agreements
* Standardized extraction protocols
* Regular data quality audits
* Proven governance structure since 2013

**MRAIA Honest Broker Services**

* ISO 27001 certified facility
* Established BAAs with all sites
* Validated deduplication algorithms

**Computing Resources:**

* **Local**: Dell PowerEdge servers with redundant storage
* **Cloud**: Microsoft Azure
* **Software**: Site licenses for R, SAS, ArcGIS, Tableau
* **Security**: Fortinet firewalls, Splunk monitoring

**Funding:**

* **Searle Family Trust**: (July 2024-July 2029)

**Training and Expertise:**

* **All staff**: CITI Human Subjects, HIPAA training
* **PI**: 7+ years of health services research, clinical informatics expertise
* **Team expertise**: Spatial analysis, health disparities, community engagement
* **Monthly meetings**: Protocol adherence and quality control

**Community Partnerships:**

**Established Relationships:**

* West Side United

**Engagement Infrastructure:**

* Community Advisory Board for **The RUSH BMO Institute for Health Equity**
* Translation services (Spanish, Polish, Mandarin)
* Policy briefs for decision makers

This comprehensive resource base ensures successful execution of this complex multi-site study while maintaining the highest standards of scientific rigor and community engagement.

1. References

1. Large life expectancy gaps in U.s. cities linked to racial & ethnic segregation by neighborhood [Internet]. NYU Langone News. [cited 2025 May 28];Available from: https://nyulangone.org/news/large-life-expectancy-gaps-us-cities-linked-racial-ethnic-segregation-neighborhood

2. Chapter 13. In: The Death Gap. University of Chicago Press; 2017.

3. Kho AN, Hynes DM, Goel S, et al. CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. J Am Med Inform Assoc 2014;21(4):607–11.

4. MRAIA [Internet]. mraia. [cited 2025 Jun 10];Available from: https://www.mraia.org/

5. Johannesen AM. Review of the death gap: How inequality kills the death gap: How inequality kills. Chicago: University of Chicago press. 240pages. ISBN: 978-0226428154. Hardcover, $26. David Ansell. 2017. Poverty Public Policy 2018;10(3):417–9.

6. Chicago Health Atlas [Internet]. Chicago Health Atlas. [cited 2025 May 28];Available from: https://chicagohealthatlas.org/

7. Home [Internet]. [cited 2025 May 28];Available from: https://www.chicago.gov/city/en/sites/vrd/home.html

8. PCORnet PPRN Consortium, Daugherty SE, Wahba S, Fleurence R. Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research. J Am Med Inform Assoc 2014;21(4):583–6.

9. Moore BJ, White S, Washington R, Coenen N, Elixhauser A. Identifying increased risk of readmission and in-hospital mortality using hospital administrative data: The AHRQ Elixhauser Comorbidity index. Med Care 2017;55(7):698–705.

10. Schuster MA, Osherov A, Chung PJ. Why counting black lives matters - the 2020 census. N Engl J Med 2020;383(9):e60.

11. Khan D, Rossen LM, Hamilton BE, He Y, Wei R, Dienes E. Hot spots, cluster detection and spatial outlier analysis of teen birth rates in the U.S., 2003-2012. Spat Spatiotemporal Epidemiol 2017;21:67–75.

12. Kramer PR. A review of some Monte Carlo simulation methods for turbulent systems. Monte Carlo Methods Appl [Internet] 2001;7(3–4). Available from: http://dx.doi.org/10.1515/mcma.2001.7.3-4.229

1. Appendices

## Appendix A: Core Chronic Disease Categories (38 Total)

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Condition Category** | **ICD-10-CM Codes** | **Validation Source** |
| 1 | Hypertension | I10-I15 | Quan et al. 2005 |
| 2 | Coronary/ischemic heart disease | I20-I25 | AHRQ CCIR |
| 3 | Congestive heart failure | I50.x | Quan et al. 2005 |
| 4 | Cerebrovascular disease & stroke | I60-I69 | AHRQ CCIR |
| 5 | Peripheral arterial disease | I70-I79 | AHRQ CCIR |
| 6 | Chronic obstructive pulmonary disease | J41-J44 | Quan et al. 2005 |
| 7 | Asthma | J45.x | AHRQ CCIR |
| 8 | Diabetes without complications | E08.9, E09.9, E10.9, E11.9, E13.9 | Quan et al. 2005 |
| 9 | Diabetes with complications | E08.2-E08.8, E09.2-E09.8, E10.2-E10.8, E11.2-E11.8, E13.2-E13.8 | Quan et al. 2005 |
| 10 | Chronic kidney disease | N18.x | AHRQ CCIR |
| 11 | Chronic liver disease & cirrhosis | K70.x, K73.x, K74.x | Quan et al. 2005 |
| 12 | Chronic viral hepatitis | B18.x | CDC |
| 13 | HIV/AIDS | B20-B24 | CDC |
| 14 | Obesity | E66.x | AHRQ CCIR |
| 15 | Hyperlipidemia | E78.x | AHRQ CCIR |
| 16 | Major depressive disorders | F32.x, F33.x | AHRQ CCIR |
| 17 | Schizophrenia & psychotic disorders | F20-F29 | AHRQ CCIR |
| 18 | Alcohol use disorder | F10.x | DSM-5 mapping |
| 19 | Opioid & other drug use disorders | F11.x, F12.x, F14-F16.x, F18.x, F19.x | DSM-5 mapping |
| 20 | Sickle cell & hemoglobinopathies | D57.x | CDC |
| 21 | Dementia including Alzheimer's | G30.x, F01.x-F03.x | Quan et al. 2005 |
| 22 | Rheumatoid arthritis & CTD | M05-M08, M30-M36 | AHRQ CCIR |
| 23 | Coagulopathy & clotting disorders | D65-D69 | Quan et al. 2005 |
| 24 | Deficiency anemias | D50-D53 | AHRQ CCIR |
| 25 | Lung cancer | C34.x | NCI |
| 26 | Breast cancer | C50.x | NCI |
| 27 | Prostate cancer | C61.x | NCI |
| 28 | Colorectal cancer | C18-C20 | NCI |
| 29 | Pancreatic cancer | C25.x | NCI |
| 30 | Liver cancer | C22.x | NCI |
| 31 | Cervical cancer | C53.x | NCI |
| 32 | Ovarian cancer | C56.x | NCI |
| 33 | Kidney cancer | C64-C65 | NCI |
| 34 | Stomach cancer | C16.x | NCI |
| 35 | Leukemias | C91-C95 | NCI |
| 36 | Lymphomas | C81-C85 | NCI |
| 37 | Multiple myeloma | C90.x | NCI |
| 38 | Sepsis/septicemia | A40.x, A41.x, R65.2 | CDC |

## Appendix B: Firearm-Related Violence ICD-10-CM Codes

**Table B1. Comprehensive Firearm Violence Codes for Surveillance**

|  |  |  |
| --- | --- | --- |
| **Category** | **ICD-10-CM Codes** | **Description** |
| **Initial Encounters (Non-Fatal)** |  |  |
| Accidental discharge | W32.0xxA - W34.9xxA | Initial encounter for accidental firearm discharge |
| Intentional self-harm | X72.xxA - X74.9xxA | Initial encounter for intentional self-harm by firearm |
| Assault | X93.xxA - X95.9xxA | Initial encounter for assault by firearm |
| Undetermined intent | Y22.xxA - Y24.9xxA | Initial encounter, firearm discharge undetermined intent |
| Legal intervention | Y35.0xxA - Y35.09xA | Initial encounter for legal intervention with firearm |
| War operations | Y36.4xxA | Initial encounter for war operations involving firearms |
| **Subsequent Encounters** |  |  |
| All categories above | 7th character 'D' | Subsequent encounter for firearm injury |
| **Sequelae (Long-term Effects)** |  |  |
| All categories above | 7th character 'S' | Sequela of firearm injury |

**Table B2. Enhanced Censoring Plan for Firearm Violence Data**

|  |  |  |
| --- | --- | --- |
| **Data Type** | **Censoring Rule** | **Rationale** |
| Initial encounters | Suppress if n<5 | Protect identity in sensitive incidents |
| Subsequent encounters | Suppress if n<10 | Less sensitive, standard threshold |
| Sequelae | Suppress if n<10 | Long-term care, standard threshold |
| Age-Sex-Race cells | Suppress if n<5 for any firearm code | Enhanced protection for demographics |
| Monthly/Quarterly data | Not reported | Annual aggregation only |
| Exact location of victim home address | Census tract only | No sub-tract geography |

## Appendix C: Data Dictionary for Aggregate Files

**Table C1. GEO\_AGGREGATES\_STRATIFIED.csv Structure**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type** | **Values** | **Description** |
| FIPS11 | Character | 11-digit code | Census tract identifier |
| YEAR | Integer | 2019-2024 | Calendar year |
| CONDITION | Character | See Appendix A + Firearm categories | Chronic disease or injury category |
| AGE\_GROUP | Character | 18-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ | Age stratification |
| SEX | Character | M, F, U | Biological sex |
| RACE | Character | WHITE, BLACK, ASIAN, AIAN, NHPI, MULTIPLE, OTHER, UNKNOWN | Race category |
| ETHNICITY | Character | HISPANIC, NON\_HISPANIC, UNKNOWN | Ethnicity |
| COUNT | Integer | 0-9999 (suppressed if <10, or <5 for firearm) | Number of patients |

**Table C2. Statistical Disclosure Control Rules**

|  |  |  |
| --- | --- | --- |
| **Rule** | **Description** | **Implementation** |
| Primary Suppression | Suppress cells with n<10 (n<5 for firearm) | Set COUNT = NULL |
| Complementary Suppression | Suppress additional cells to prevent back-calculation | Algorithm selects minimum cells |
| Total Suppression | If >30% cells suppressed, suppress entire stratum | Protects against sparse data |
| Geographic Aggregation | Combine adjacent tracts if needed | Only for tracts with <1000 population |
| Firearm-Specific Rules | No exact counts 1-4, no sub-annual reporting | Additional privacy protection |

**Appendix D: Geographic Coverage Detail**

**Table B1. Illinois Counties Included in CONSCIENCE Study**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **County** | **2024 Population** | **Census Tracts** | **Major Municipalities** | **FIPS Prefix** |
| Cook | 5,173,000 | 1,319 | Chicago, Evanston, Oak Park, Cicero | 17031 |
| DuPage | 932,000 | 219 | Naperville, Aurora (partial), Wheaton | 17043 |
| Lake | 714,000 | 176 | Waukegan, Lake Forest, Highland Park | 17097 |
| Will | 696,000 | 168 | Joliet, Bolingbrook, Orland Park | 17197 |
| Kane | 532,000 | 121 | Aurora (partial), Elgin, St. Charles | 17089 |
| McHenry | 310,000 | 77 | Crystal Lake, Woodstock, McHenry | 17111 |
| **Total** | **8,357,000** | **2,080** | **200+ municipalities** | — |

**Appendix C: Data Dictionary for Aggregate Files**

**Table C1. GEO\_AGGREGATES\_STRATIFIED.csv Structure**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type** | **Values** | **Description** |
| FIPS11 | Character | 11-digit code | Census tract identifier |
| YEAR | Integer | 2019-2024 | Calendar year |
| CONDITION | Character | See Appendix A | Chronic disease category |
| AGE\_GROUP | Character | 18-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ | Age stratification |
| SEX | Character | M, F, U | Biological sex |
| RACE | Character | WHITE, BLACK, ASIAN, AIAN, NHPI, MULTIPLE, OTHER, UNKNOWN | Race category |
| ETHNICITY | Character | HISPANIC, NON\_HISPANIC, UNKNOWN | Ethnicity |
| COUNT | Integer | 0-99999 (suppressed if <10) | Number of patients |

**Table C2. Statistical Disclosure Control Rules**

|  |  |  |
| --- | --- | --- |
| **Rule** | **Description** | **Implementation** |
| Primary Suppression | Suppress cells with n<10 | Set COUNT = NULL |
| Complementary Suppression | Suppress additional cells to prevent back-calculation | Algorithm selects minimum cells |
| Total Suppression | If >30% cells suppressed, suppress entire stratum | Protects against sparse data |
| Geographic Aggregation | Combine adjacent tracts if needed | Only for tracts with <1000 population |