Measuring New York City: A Cautionary Tale

Amanda Geller
Associate Research Scientist
Columbia University Schools of Social Work and Law
Columbia Population Research Center

July 15, 2011

Address all correspondence to:

Amanda Geller
Columbia Population Research Center
1255 Amsterdam Avenue
New York, NY 10027
Tel. 212-851-2380
Email: abg2108@columbia.edu
ABSTRACT

This paper identifies sources of data relevant for the analysis of New York City crime trends, and for contextualizing crime patterns in order to assess their causes and effects. Focusing on the twenty-year period from 1990-2009, I review methodological and substantive concerns related to the study of crime trends, with particular attention to the geographic and organizational structure of New York City. I then provide an inventory of data sources related to criminal activity in New York City, and to the social, economic, and law enforcement factors that provide context for understanding changes in local crime patterns. While numerous data sources exist at the City, State, and National levels, the variability in data quality is noteworthy. Few data sources span the full two decades of the crime drop, and those that do tend to focus on large geographic areas such as boroughs and the city as a whole. Data reflecting greater geographic detail tend to be available for shorter time horizons. Some data sources can be combined to span the time period of interest; however, care must be taken in doing so. Data sources that may be presumed to measure the same phenomena do not necessarily align, and little documentation is publicly available to reconcile differences. While researchers may impute or interpolate missing data, allocate data across geographic boundaries, or append consecutive time series, the resulting empirical analyses will require a variety of robustness tests to ensure that research findings are not sensitive to variations in data assumptions.

ACKNOWLEDGMENTS

This review could not have been written without the cooperation of several colleagues and staff members at supporting organizations. I am grateful for the assistance of Jeffrey Fagan, Frank Zimring, Freda Solomon, Gina Lovasi, Andrew Case, James Quinn, Jeffrey Lane, Christine Bruzzese (City Hall Library), Brian Smidt (National Insurance Crime Bureau), Mark D. Miller (National Insurance Crime Bureau), Michael Bauer (NYS Dept. of Health), Catherine Stayton (NYC DOHMH), Renee Konicki (DCJS), Marge Cohen (DCJS), and Christopher Dunn (NYCLU). All errors and conclusions are my own.
Measuring New York City: A Cautionary Tale

I. Introduction

After historically high crime rates in the 1980s, New York City experienced a sharp, unexpected, and unprecedented crime decline\(^1\). The magnitude and unexpected nature of the decline has inspired a range of research questions, related to the causes of the crime reduction, the role of policing in maintaining low crime rates, and the consequences of crime and criminal justice policy for public safety\(^2\). Empirical research on the causes and consequences of crime in New York is highly dependent on the quality of available data, and requires a thorough understanding both of data sources and of a variety of analytical issues.

This paper identifies sources of data that are relevant for both the measurement of New York City crime trends, and for contextualizing crime patterns in order to assess their causes and consequences. I begin by reviewing methodological concerns of general relevance to the study of crime trends, and measurement concerns particular to the study of New York City. I then provide an inventory of sources that provide data related to criminal activity in New York City. I also provide an inventory of data sources related to social, economic, and law enforcement factors that provide context for understanding changes in New York City crime patterns. Finally, I provide a summary of data availability and quality, and suggest strategies for combining data sources to use in a comprehensive analysis.

II. Conceptual Issues

---


A. Measurement

When discussing and analyzing a “crime decline”, it is important to bear in mind that crime trends are based on individual incidents, which each take place in a singular location at a point in time. Empirical research on these trends thus depends on the reliability of incident reporting, and the aggregation of data over space and time. When measuring crime and social trends, and particularly when comparing data reported by criminal justice authorities to those reported by other sources (e.g., city and state departments of health), it is important to note that incident reports may have different substantive meanings, depending on the reporting source. For example, “property crime” may encompass incidents of varying degrees of seriousness, depending on whether the designation refers to felonies, misdemeanors, or a combination of the two. Even designations such as “assault” and “homicide” take on different meanings in the health and criminal justice communities. It is therefore logical that data tabulations may not directly correspond directly across sources; the treatment of disparate sources must be appropriate for the research question of interest.

The appropriate level of either spatial or time aggregation is also dependent on the research question of interest. Given the stochastic nature of crime, day-to-day fluctuations are expected. However, crime rates in small time units are likely to be autocorrelated, and resemble a “random walk” where the city’s crime “position” in a given day is so closely tied that of the previous day that changes might not be meaningful. On the other hand, units that are too large might mask meaningful variation and trends. Likewise, the factors influencing crime patterns can be measured at a variety of spatial levels, based both on local social structure (e.g., “hot spots”) and administrative concerns (e.g., citywide programs, or precinct commander priorities).

While units that are too small may be influenced by the spillover of social phenomena between adjacent areas, units that are too large may mask meaningful heterogeneity. For the purposes of the data source inventory that follows, sources will be reported based on the smallest spatial and time units at which data are available, as these units can be aggregated as researchers deem appropriate.

B. Geography

In this section, I review the types of areal units most commonly used when analyzing crime trends and social structure in New York City. In addition to the City’s five boroughs (The Bronx/Bronx County, Brooklyn/Kings County, Manhattan/New York County, Queens/Queens County, and Staten Island/Richmond County), I describe units used by the NYPD, and then describe units commonly tracked by the U.S. Bureau of the Census, and other geographic units by which social phenomena in New York are commonly measured. Areal units that are mutually exclusive and span the entire city are noted in **bold**.

1. NYPD Units

The NYPD has organized the city into eight **patrol boroughs**, and 76 police **precincts**. A precinct map of the city is provided in Appendix A. Precincts are the unit at which most police activity is assigned and regulated; uniformed officers are organized into these 76 precincts, with additional command units dedicated to the subway system, public housing

---


6 Manhattan South, Manhattan North, Brooklyn South, Brooklyn North, Queens South, Queens North, Bronx, and Staten Island.

7 The Precinct structure has been constant since 1994, when the 34th precinct split from the 33rd into its own distinct unit. It is notable that this structural change came in the time period of the proposed analysis.

8 However, precinct activities are also supervised by the Patrol Services Bureau (PSB).
projects, schools, and other sites of interest. The NYPD’s Transit Bureau has 12 “Transit Districts” throughout the four largest boroughs\(^9\), and the Housing Bureau has nine “Housing Police Service Areas (PSAs)”. Transit and Housing units geographically overlap with precincts, and crime and police activity reported to housing and transit units are recorded with the precinct in which they take place.

Police precincts are further subdivided into \textbf{beats} and \textbf{sectors}. Another geographic division of note is driven by Operation Impact, introduced by the NYPD in 2003, which aims to increase police presence in strategically targeted locations known as “Impact Zones”\(^10\). Impact zones are approximately the size of Census Tracts\(^11\), described below, but vary over time, are not systematically determined, and may span precinct boundaries\(^12\).

2. Census Tracts and Aggregated Units

While police activity, including the reporting of crime, is largely organized by precinct, New York City social structure is generally reported at other geographic levels. As the organization of record for demographic and socioeconomic information, the Census Bureau relies on a number of geographic units\(^13\), which are classified as either “legal and administrative” (i.e., developed from legal actions, treaties, ordinances, etc.) or “statistical” (i.e., developed by the Census Bureau as a result of “practice, custom, usage, or need.”) In discussing crime and its covariates in New York City, the smallest areal unit of interest is likely to be the Census

---

\(^9\) An approximate list of the subway stations covered by each transit bureau is available at [http://www.nyc.gov/html/nypd/html/transit_bureau/transit.shtml](http://www.nyc.gov/html/nypd/html/transit_bureau/transit.shtml); however, these allocations do not reflect recent changes to the MTA system (or even older ones such as the removal of the 9 train), and contain several errors (e.g., the C train does not go to the Bronx).


\(^13\) See [http://www.census.gov/geo/www/tiger/glossry2.html](http://www.census.gov/geo/www/tiger/glossry2.html) for an exhaustive list.
Tracts typically have populations between 1,200 and 8,000 people, with an “optimum size” of 4,000 people. As of the 2000 Census, the City had 2,217 tracts. It is notable, however, that the population of New York City tracts varies well beyond the range generally specified; the City has tracts with fewer than 500 residents, and tracts with more than 10,000. (The largest tract, Co-Op City in the Bronx, has more than 24,000 residents).

Tracts may be aggregated into a number of units of potential interest. For example, the New York City Department of City Planning had previously divided the city into 292 neighborhoods, but currently divides the city into 195. The New York City Housing and Vacancy Survey divides the city into 55 subboroughs, which coincide with the Census Bureau’s Public Use Microdata Areas (PUMAs).

Notably, while Census tracts are exact subdivisions of counties, PUMAs/subboroughs, and neighborhoods, the NYPD’s units do not coincide with Census units. Tracts may span precinct boundaries, complicating the computation of precinct demographic and socioeconomic data.

---

14 Census Tracts are defined as “a small, relatively permanent statistical subdivision of a county or equivalent entity… that are updated by local participants… to provide a stable set of geographic units for the presentation of statistical data.” (See: http://www.census.gov/geo/www/2010census/gtc_10.html, hereafter “Census, 2010”). Tract boundaries tend follow visible and identifiable features (e.g., rivers, highways, etc.), and state and county boundaries will always coincide with Census tract boundaries (suggesting also that tract-level data can be aggregated to either the county or state level). While defined as “relatively permanent”, tracts can be split or merged from one census to the next, based on population changes. This is a rare, but noteworthy, occurrence.
15 The smallest areal unit defined by the Census Bureau is the Census Block, which is defined as “statistical areas bounded by visible features, such as streets, roads, streams, and railroad tracks, and by nonvisible boundaries, such as selected property lines and city, township, school district, and county limits and short line-of-sight extensions of streets and roads.” and is the basis for all tabulated data. However, blocks and Block Groups, while subdivisions of Census Tracts, are generally too small to be of interest for data analysis, since most data of interest to social scientists is reported at larger levels of aggregation.
18 The old boundaries are still used by some reference services, such as Infoshare.
20 PUMAs are geographic areas for which the Census Bureau provides selected extracts of raw data from a sample of Census participants (Census, 2010, supra note 14).
characteristics. Precinct demography and socioeconomics may, however, be estimated by allocating tract characteristics to precincts based on the extent to which their areas overlap.

3. Other units

As with police precincts, several other geographic units of interest do not perfectly coincide with the boundaries of Census Tracts. In addition to the ZIP codes used by the U.S. Postal Service, the City is divided into 59 **Community Districts (CDs)**, and each district profile contains summary demographic, socioeconomic, and land use data. CDs are qualitatively similar to **PUMAs**, though not coterminous\(^{21}\), and while PUMAs can be defined by the tracts they contain, CDs cannot, as tracts may cross CD boundaries. Since 1927, the City Department of Health and Mental Hygiene has also divided the city into 30 **Health Center Districts (HCDs)** to report mortality data. While these are given “neighborhood names”, HCDs do not correspond to CD’s or the City Planning neighborhoods. The other unit of note for reporting health data is the United Health Fund (**UHF neighborhood**. While the number of UHF neighborhoods used by the City has varied over time, they tend to consist of between three and six ZIP codes, and reflect catchment areas for healthcare facilities\(^{22}\).

The City is also subdivided into several other sets of administrative units, such as **school districts** and **Election Districts** (including **City Council Districts, State Senate Districts, State Assembly Districts**, and others), which do not systematically correspond to either NYPD or Census units\(^{23}\). However, none of the data sources listed below utilize these units, so they are likely to be of limited interest for the current analysis.

---

\(^{21}\) See, for example: http://www.nyc.gov/html/dcp/pdf/census/puma_cd_map.pdf


III. Crime Data

A. Crime Count Data

Since the 1980s, crime data has been available from several different sources, each for years at a time. However, no single authoritative source has provided crime data consistently from 1990-2010. In the sections that follow, I therefore lay out the variety of sources in which data are available, and discuss concerns for reconciling multiple sources.

1. “Arrest and Complaints” (hereafter “AC”) data

AC data were reported annually for each precinct from 1984-2000, and monthly for each precinct from 1989-2001. For each precinct in each month, the number of arrests and complaints made for the seven major offense categories (murder/non-negligent homicide, forcible rape, robbery, assault, burglary, grand larceny, and grand larceny-auto), as well as categories for “other felonies”, misdemeanor reports and arrests, misdemeanor “youth” reports, misdemeanor “summ” reports, violation reports and arrests, and violation “youth” and “sum” reports. Arrest data are reported with age breakouts (Under 16/16-20/21-24/25+) for felony arrests, misdemeanor arrests, and violation arrests.

While the AC data are valuable for precinct-level totals, significant data is lost in aggregation. Racial breakdowns are not reported for either arrestees or crime suspects, and no information is provided about drugs or weapons, salient points in the NYPD’s stated and observed enforcement strategies. It is also notable that if arrests do not correspond to a specific


25 Provided by the NYPD.
crime complaint, a complaint report is generated to correspond to the arrest. It is difficult, if not impossible, to distinguish arrest-generated complaints from others in aggregate data.

2. Compstat

One of the most significant management changes in policing in the past 20 years has been the implementation of Compstat, an accounting and analysis tool that generates data for systematic analysis of location-specific crime trends. This location-specific data allows police commanders to rapidly allocate resources in response to changes in crime patterns. Compstat also tracks detailed data on arrest activity, often used to assess officer performance. Incident-level data are closely protected by the NYPD and are not widely available.

Aggregate data from Compstat, however, are publicly available, and updated weekly on the NYPD’s website. While cumulative records are not maintained online, back-issues of Compstat reports from 2000-2008 are available in print at the City Hall Library. Annual totals are also available in the “Year End Reports”, which cover the week from 12/25-12/31 of each year, and contain “year to date” information. Weekly Compstat reports contain the number of complaints reported for each of the seven major crime types (murder, rape, robbery, felony assault, burglary, grand larceny, grand larceny – auto, and “other”), both precinct-wide, and split between units overseen by each Patrol Services Bureau (PSB), transit commands, and housing.

26 See http://www.nyc.gov/html/nypd/html/analysis_and_planning/historical_nyc_crime_data.shtml: “An arrest that is proactive in nature (i.e. drug arrest) is counted within the presented statistics as a crime complaint in the appropriate category.”


29 31 Chambers Street, Room 112, NY, NY 10007. Notably, early years’ Compstat reports are much less frequently available (only a few weeks from 2000 and 2001 are available, though the archives seem to be more complete from mid-2002 through 2008.) The Archive does not have Compstat data after 2008, and is unlikely to get them.
commands. In addition, several complaint types of particular interest are separately identified: shooting victims, shooting incidents, and Rape 1. Archived Compstat reports also contain the number of arrests made in each precinct in each week, in the seven major offense categories, and for other offenses of interest (e.g., “Gun”, “Narcotics”, and violations of Vehicle Traffic Law 511 – Driving with a suspended license). In addition to PSB, transit and housing officers, these also track arrests made by the Organized Crime Control Bureau (OCCB) and Detective Bureau.

While the City Hall Library contains tremendous amounts of underutilized data, data collection efforts should proceed with caution. The City Hall Library has no additional information from the NYPD, such as explanations for acronyms or data collection procedures, which might help in interpretation of the Compstat data. In addition, although complaints for the seven major felony offenses appear to be tracked consistently over time, other data published in Compstat reports tends to vary. For example, arrest information is no longer tracked; nor is information related to summons or warrant activity. As with the AC data, crime complaints generated by an arrest are difficult to distinguish from those reported by victims or witnesses.

Also lacking with respect to the Compstat data is a precise formula for reconciling Compstat reports of arrests and complaints with reports in earlier years and from other sources. An analysis of one year reported in both the AC file and the Compstat file (2001) suggests that precinct annual totals are not directly comparable between data sources, and, as noted, no documentation is provided of changes in data collection protocol. Transitions between the earlier and later time series must be handled carefully, and include a wide range of several sensitivity analyses.

3. Furman Center


The Furman Center at New York University has a “data search tool” that includes annual crime totals at the precinct level, obtained from the NYPD. However, the extent to which the Furman Center data matches other NYPD sources is unclear. Spot checks of 2001 data (also reported in the AC and Compstat files) suggest correspondence with the AC file on murder, robbery, and rape counts, but differences on other indicators, including substantial differences on assaults, burglaries, and larcenies. Differences also exist between the Furman Center and Compstat files.

4. Infoshare

Infoshare, described in greater detail below, is a subscription-based data archive run by Community Studies of New York, Inc., a non-profit corporation based in New York City. Infoshare provides annual crime data at the Community District Level from 1985 through 2001. Data include reported felonies (7 major offenses itemized, plus “other felonies”), as well as felony, misdemeanor, and violation arrests, both CD-wide, and by age of arrestee (under 16, 16-20, 21-24, and 25+). Data were provided to Infoshare by the NYPD.

5. NYPD Website

In addition, the NYPD’s official website contains a page of “Historical New York City Crime Data”\(^\text{32}\). This page contains citywide annual totals of complaints for the Seven Major Felonies, Other Felony Crime, Misdemeanors, and Violations, by individual crime types, between the years 2000 and 2010. Offense descriptions for which “a substantial portion of the complaint counts are generated as a result of the implementation of pro-active policing strategies” are noted, and include Felony Possession of Stolen Property, Felony (and Misdemeanor) Dangerous Drugs, Felony (and Misdemeanor) Dangerous Weapons, Criminal Trespass, and several others.

Although the NYC.gov data contain substantially more detail about the types of offenses reported to the police between 2000 and 2010, important information is lacking. First, spot check comparisons of the 2000 data with AC totals, and the 2001 data with Compstat totals, suggest close, but not exact, matches between the data sources. The source of observed mismatch is unclear. The NYC.gov data also do not contain any information about the geographic distribution of crime, or variation over the course of the year (e.g., months or calendar quarters). Accordingly, these data will be insufficient to determine whether local policing efforts have contributed to changes in crime levels. The NYC.gov data also say little about arrests made, other than to indicate that several types of crime complaints are driven heavily by policing activity (and, presumably, arrests).

B. Proxy Sources

Although crime statistics are, as noted, available in several forms from City and criminal justice agencies, the accuracy of police statistics has been called into question\(^\text{33}\). Criminal justice scholars have therefore worked to verify official statistics with data from independent sources, such as victim surveys and health statistics\(^\text{34}\). Although no comprehensive victim surveys allow city-level evaluation of changes in crime levels over time\(^\text{35}\), some verification is possible from data from the Department of Health, and from insurance statistics.

1. Health Statistics

The City’s Vital Statistics (VS) program\(^\text{36}\) contains publicly available data on the number of deaths caused by homicide from 1961 through 2009, both citywide and by decedent

---


\(^{34}\) Zimring, forthcoming, supra note 1.

\(^{35}\) ibid.

residence. Zimring (forthcoming) finds close correspondence between NYPD and DOH citywide homicide statistics. Health data can also be used to supplement the minimal NYPD data available related to drug crime and weapons possession. Borough-level data are available from 1985-2009 for mortality related to “accidental poisoning by psychoactive substance”. VS data may also be useful in the assessment of firearms trends. Annual VS reports contain information on the number of deaths by firearm across the city from 2000-2009, as well as, from 1980-2009, the number of suicides and homicides that involved a firearm. Data on accidental deaths involving firearms are available from 1990-2009.

State-level data from the Statewide Planning and Research Cooperative System (SPARCS) also contain data on hospitalizations from 1996 through current years. While New York City also collects some data on hospitalizations, this surveillance is done only on a small scale. The New York City DOHMH is an authorized user of the SPARCS system, and relies on the state data for systematic tracking. New York City hospitals can be identified as a single “Health Service Area”, or by region (Bronx County, Kings County, Queens and Richmond County combined, and two regions within New York County). Hospitalizations are recorded by CDC “E-codes”, which contain classifications such as “accidental poisoning from psychotropic agents”, “firearm”, “cut/pierce”, and others, with special codes to identify injuries resulting from

---

38 Accidental deaths involving firearms are not reported in 2005.
39 Stayton, Catherine, NYC DOHMH. 2011. Personal communication, 7/13/11.
40 SPARCS data are not publicly available, but can be requested as specified at http://www.health.ny.gov/statistics/sparcs/forms/non-id.htm
assaults\textsuperscript{41}. Emergency department data was not collected by the state DOH until 2005, and are likely of limited use in assessing long-term trends\textsuperscript{42}.

2. Insurance Data

Zimring (forthcoming) also compares trends in NYPD-reported theft with data from the National Insurance Crime Bureau (NICB), and has made this data available for further analysis. The NICB data run from 1986 to December 2009, and contain the number of thefts reported by NICB member insurance companies. Data are reported at three levels of detail: MSA, New York City alone, and ZIP code within New York City. However, the ZIP code-level data are based not on the area where thefts occurred, but on the location of the reporting agency. The NYPD’s stolen property bureau is located in ZIP code 10038; nearly all thefts are thus associated with the 10038 ZIP code, and do not reflect the true distribution of thefts in New York City.

C. New York State Division of Criminal Justice Services (DCJS): Additional Information

As a clearinghouse of criminal justice data for New York State, the DCJS contains additional information that might be valuable for assessing crime trends in New York City. For example, the DCJS tracks the number and percent of murders, rapes, robberies, and aggravated assaults committed with firearms, with data available from 1990 through 2002. These data are generally available at the City level, and thus cannot be used for assessing the geographic distribution of crime within the city. It is also notable that the NYPD stopped providing portions of this data in 2002; the statistics can thus only be used to assess the first part of the crime drop.

DCJS may also be useful for providing data about arrestees, both as an aggregate approximation of crime trends and in terms of assessing individual trajectories as arrestees negotiate the criminal justice system. Arrest data are drawn from the DCJS’s Computerized

\textsuperscript{41} However, as noted above, “assault” designations may not correspond to penal law designation.

\textsuperscript{42} Furthermore, the 2005 and 2006 ED data may suffer from some startup challenges, so trend analyses should begin with 2007. (Stayton, Catherine, NYC DOHMH, Personal Communication, 7/13/11).
Criminal History System, and contain information on the precinct and month in which individuals were arrested, and the crimes they were charged with. Race/ethnicity data is available beginning in 2003\(^43\). The release of individual-level case files is documented in the DCJS Data Request Policy and requires a detailed data request process, detailed in Appendix B. Larger aggregations of data (e.g., city-level annual arrest totals) can be requested without a detailed nondisclosure process. Aggregations at more intermediate levels (e.g., totals by year and charge) may require the more detailed request process, depending on the particular information of interest.

IV. Crime Covariate Data

In order to better understand the causes and correlates of the crime decline, it is necessary to consider the local context in which criminal activity takes place. In a 1993 NAS report on violence\(^44\), Sampson and Lauritsen discuss six predictors of community crime that may be of interest while analyzing New York City: (1) Poverty and deprivation, (2) Mobility and community change, (3) Housing and population density, (4) Family structure, (5) Community social disorganization [or organization], and (6) Community cultures and ethnography. This paper adds a seventh set of covariates: criminal justice system factors. However, each of these constructs may present challenges for measurement. This section therefore begins with a discussion of each construct and measurement issues involved, followed by a discussion of covariate data sources, each of which may span several substantive domains.

A. Covariates of Community Crime

\(^43\) The DCJS instituted new standards for their race/ethnicity data in 2003, and does not provide data that does not meet (ie, preceded) these standards. (Cohen, Marge, DCJS. 2011. Personal Communication.)

1. Poverty and deprivation

Local economic wellbeing and deprivation may be measured in several different ways. Poverty rates (e.g., the percent of the area population living in poverty) can be used to measure absolute levels of deprivation, while population employment and unemployment rates can be used as another economic indicator. However, it must be noted that current measures of poverty are insufficient to reflect the nuances of socioeconomic conditions. For example, a binary indicator of whether a person or household is “in poverty” not only assumes that the poverty line accurately reflects their income and needs, but also fails to identify individuals in deep poverty or the implications of such disadvantage. Furthermore, to the extent that geographic areas are heterogeneous, unitwide poverty rates may mask smaller areas of concentrated poverty. Most aggregate data sources are limited in their ability to measure local income distributions; however, some measures of economic inequality can be estimated from aggregate sources.

2. Mobility and community change

Indicators of neighborhood change can be ascertained in two ways. Respondents in cross-sectional surveys can retrospectively report how long they have lived in their current residence or neighborhood; however, retrospective reports are vulnerable to recall error, and should be interpreted with caution. Alternatively, changes in community conditions can be measured with longitudinal data collection.

3. Housing and population density

---


46 For example, when income is reported by the number of households within a given range (e.g., <$10,000, $10,000-14,999, $15,000-24,999, … , $200,000 or more, income quantiles (10th percentile, 20th percentile, median, etc.) can be approximated based on the number of households in each range. Using these approximate income quantiles, income disparities can be estimated using a 90:10 ratio, 80:20 ratio, 50:10, etc.
Population density can be computed from basic population and land area measures. Other housing and land features, such as home ownership, the built environment and land use, may also have implications for crime trends. In addition, public housing may have unique and relevant features for criminal and policing activity. Public housing can be identified in neighborhoods or other areal units in several ways: the presence of projects, the land area devoted to public housing, or the population living in public housing; however, it is notable that administrative data on public housing residents may undercount those not on household rosters.

4. Family structure

Family structure may predict criminal activity either as an indicator of socioeconomic status, or as a proxy for parental supervision that might prevent juvenile offending. However, family structure may be difficult to measure in citywide surveys or administrative data, as large-scale data collection may not capture the presence of grandparents, nonmarital partners, or other nonparental caregivers, or indicators of family instability, such as relationship transitions or multiple partner fertility. Official measures of family structure are therefore likely to only approximate the conditions that might enable or prevent illegal activity.

5. Community social disorganization [or organization]

While each of the demographic and socioeconomic factors described thus far contribute to community social structure, other local conditions may also provide unique insight into local social organization. For example, in addition to the employment and unemployment rates described above, which measure the employment status of local residents, regardless of where they work) and an aspect of neighborhood socioeconomic status, data on local employment (e.g., the number of people working in a neighborhood, regardless of where they reside) may provide a

---

different measure of local social structure. Likewise, land use factors, such as the presence of Business Improvement Districts (BIDs) may have implications for local crime\textsuperscript{48}. Data on community schools may also be of interest; while criminologists may not be well-equipped to measure school quality, data on absences, suspensions, or school closures may be of interest as a proxy for social disorganization. However, it is also important to note that while changes in these measures may reflect changes in student behavior; they may also reflect changes in the leniency or stringency of assessment.

6. Community cultures and ethnography

In a city as diverse as New York, and given the racial skew of the population under correctional supervision\textsuperscript{49}, data on race and ethnicity are important components of any criminal justice analysis. However, the substantive meaning of race/ethnic data is highly dependent on the quality of available data, and on the reporting source. For example, the majority of New York City crime complaints in recent years were reported without a suspect race description\textsuperscript{50}; conclusions drawn from the cases with suspect race recorded may not be representative of crime trends more broadly. In addition, self-reports of race/ethnicity may differ from reports provided by others. For example, NYPD data suggests that 30\% of street stops made in Washington Heights and Inwood (i.e., 33\textsuperscript{rd} and 34\textsuperscript{th} precincts) between 2008 and 2009 were of “black Hispanics”, and 44\% were of “white Hispanics”. However, Census data suggest that nearly 60\% of area residents identify themselves as Hispanics of “other race”, a category not listed in the NYPD data, while only 15\% identify as “white Hispanics” and 4\% identify as “black Hispanics”.


\footnotesize{\textsuperscript{49} Western, Bruce. 2006. \textit{Punishment and Inequality in America}. New York, NY: Russell Sage Foundation.}

\footnotesize{\textsuperscript{50} Author’s calculations.}
The extent to which stop activity reflects the neighborhood population is thus hard to determine, but underscores the potential need to reconcile varying data sources.

Given the high rates of immigration over the past twenty years, data on immigrant populations are also important for studying the crime drop and other city conditions. However, it must be noted that survey data may underreport immigrant populations, particularly undocumented City residents. When discussing immigration in New York City, analytical decisions must also be made about the treatment of Puerto Ricans, particularly those born on the island, a U.S. territory. Puerto Ricans are United States citizens and therefore not subject to the entry restrictions of foreign citizens; however, Puerto Rican neighborhoods may resemble other immigrant neighborhoods in the presence of newcomers to the mainland United States, and the dominance of languages other than English.

With these caveats in mind, demographic data may be of interest both to determine the composition of city localities, and to compute measures of racial fragmentation and segregation. Notably, segregation measures generally require data at multiple levels of aggregation.

7. Criminal justice system and other factors

In addition to the social and demographic factors cited by Sampson and Lauritsen, police activity and other aspects of the criminal justice system are critical to consider as potential causes and effects of changes in city crime rates, as are legal, but high-risk behaviors (e.g., alcohol consumption). However, in each of these measures, the quality of data must be carefully considered, with limitations noted. One example of data quality issues with substantive implications centers around Stop, Question, and Frisk (SQF) activity and its documentation. Despite the 2003 Daniels consent decree requiring data disclosure, data were unavailable until a

---

FOIL lawsuit by the New York Civil Liberties Union in 2008. It is acknowledged by supporters and critics of the practice alike that SQF activity, reported on “UF-250 forms” has been historically underreported. While reporting is thought to have improved drastically since Daniels, reflected in rising stop counts throughout the mid-2000’s, a portion of SQF activity may remain unreported; this is difficult to determine. To the extent that respondents are likely to underreport stigmatized behaviors, survey data on risk behaviors, and survey data more generally, must also be interpreted with caution.

B. Data Sources for Crime Covariates

Many of the data sources that contain demographic, social, and economic information on New York City are cross-cutting and cover multiple substantive topics. This review lays out sources of social indicators, generally organized by reporting agent, and describes the both the type of information available and any caveats that should be considered when using the data.

1. Survey data

Survey data presents a tradeoff: small-scale surveys such as the City’s Youth Risk Behavior Survey contain detailed information about substantive areas of interest, but lack the sample size or repeated measures to be representative of small geographic areas, or of a long time horizon. Larger surveys, particularly the decennial Census, may be representative of areas as small as tracts and block groups, but tend to be conducted infrequently or lack the substantive detail of smaller and more specialized surveys. Nonetheless, both types of surveys have value for studying the causes and correlates of the long-term changes in crime patterns.

---

52 New York Civil Liberties Union v. New York City Police Department, Index No. 07/115154, 2008 WL 2522233 (New York County Supreme Court, May 7, 2008).
53 CITE
54 In addition to stops that are unreported altogether, the UF-250s may undercount stop activity in cases that stops lead to arrests, and the officers complete arrest paperwork but no UF-250.
a. U.S. Census Bureau

The leading source of data about the nation’s population and economy is the United States Census Bureau, which provides several products of interest.

i. Decennial Census

Every ten years the Census Bureau conducts the Population and Housing Census (i.e., the decennial census), which attempts to count every United States resident and collect basic demographic and socioeconomic data about them, their families, and their households. Census data may be aggregated to the tract level, and allocated to police precincts and other units. The 1990, 2000, and 2010 Censuses each contain a short form, sent to every U.S. household, which asks questions of: (1) Household size, (2) Home ownership/renter status, and (3) Age, gender, and race/ethnicity of each resident. Responses are reported in Census Summary Files 1 and 2.

The 1990 and 2000 Censuses also contain a long form, sent to a 1-in-6 sample of households, with responses reported in Summary Files 3 and 4. The long form contains more detailed population and housing data. The 2010 Census does not contain a long form, having replaced the long form with the annual American Community Survey.

ii. American Community Survey (ACS)

The ACS is an ongoing survey that collects data from a sample of Americans every year, on topics related to demography and family relationships, education, income, health insurance, where they work, and where they live. Because the ACS collects data from a sample each year, descriptive statistics that it produces are estimates that may not be representative of small social areas in each year. However, data may be pooled across years to provide a larger sample; estimates from multi-year periods are representative of smaller geographic areas. In other words,

---

56 e.g., place of birth, education, employment status, income, value of housing unit, year structure built, tenure in residence
ACS one-year estimates are the most current, released in the year immediately following the year in which they are collected; however, they are representative only of areas containing 65,000 population or more. Three-year estimates are representative of areas containing 20,000 population or more, while five-year estimates are representative of areas as small as census tracts and block groups. The first three-year estimates were released in 2008 (based on data collected from 2005 through 2007), and the first five-year estimates were released in 2010 (based on data collected from 2005 through 2009). Three and five year estimates are updated annually by removing the earliest year and replacing it with the latest one, and will provide the ability to monitor social and economic trends in local communities. Although the questions asked in the ACS are very similar to those asked in the 2000 Census Long Form, statistics are not necessarily comparable between the two data sources. The Census Bureau details important differences.

iii. Journey to Work Files

Since the 1980 Census, the Census has also contained “journey to work” questions, in which respondents report their place of employment and their place of residence. The 1980 and 1990 Censuses estimate workplace information at the county level, and are of limited use for assessing smaller social areas. However, the 2000 Journey file contains data at the tract level, and has the potential to identify “high employment nodes” where the social structure may differ from more residential neighborhoods. The ACS also contains questions about workplace and may be used to identify high employment areas and changes over time.

b. New York City Surveys

---

In addition to the national sources detailed above, several surveys specific to New York City may be of interest.

i. New York City Housing and Vacancy Survey (NYCHVS)

The NYCHVS, sponsored by the Census Bureau and the NYC Department of Housing Preservation and Development, is conducted approximately every three years, and collects data on home ownership and rentals, the rental vacancy rate, and several indicators of residential building structures, housing quality, and neighborhood conditions. Respondent households are identified by their sub-borough; however, the Census Bureau notes that “the sample size for sub-borough areas is small”, and that “comparisons of sub-borough areas should be done with great caution.” With these caveats in mind, however, questions related to housing and neighborhood conditions may be of interest when assessing the implications of physical disorder for crime.

ii. New York City Youth Risk Behavior Survey

The New York City Youth Risk Behavior Survey has been conducted through an ongoing collaboration between the New York City Department of Health and Mental Hygiene, the Department of Education, and the National Center for Disease Control and Prevention. The survey has been conducted in odd-numbered years since 1997; public school students complete a self-administered, anonymous questionnaire that measures a variety of risk behaviors, including tobacco, alcohol and drug use, unintentional injury and violence, and sexual behaviors, among others. The results are representative of public high school students across the city, in each of the five boroughs, and in three high-risk areas: the South Bronx, North and Central Brooklyn, and East and Central Harlem. Data are also available by age, sex, years in the United States, and

---

60 See: http://www.census.gov/hhes/www/housing/nychvs/2008/nychvs08.html
race/ethnicity, as well as by borough and the high-risk areas listed above. Raw data may also potentially be accessed by directly contacting the sponsoring agencies.\textsuperscript{62}

iii. New York City Community Health Survey (CHS)

The New York City CHS is a telephone survey conducted annually since 2002 by the City Department of Health and Mental Hygiene, the Division of Epidemiology, and the Bureau of Epidemiology Services. The CHS is a cross-sectional survey that contain information on risk behaviors such as binge drinking and drug use, and limited information on behaviors related to violence and firearms. Approximately 10,000 adults are surveyed each year, and estimates are representative at the borough level.

2. Administrative data

In addition to the population statistics available from survey data, community conditions may also be measured using data from city, state, and national agencies.

a. NYPD

Following the \textit{NYCLU v. New York City Police Department} decision, the New York City Police Department has made microdata on Stop, Question, and Frisk (SQF) activity publicly available online.\textsuperscript{63} Each stop reported on a UF-250 from 2003-2010 is cataloged in a database, indexed by precinct, year, and “serial number”, and containing information on the individual stopped (their age, race/ethnicity, and physical description), the suspected crime that generated the stop (as well as a series of binary indicators of several stop justifications, and if relevant, frisk and search justifications), and notable stop outcomes (including, among others, arrests, summonses, frisks, searches, seizures, use of force). Notably, SQF activity in each year is tracked in a separate file, and file formats may change slightly from one year to the next. In

\textsuperscript{62} Email to: \texttt{query@health.nyc.gov}
addition to differences noted in the NYPD’s file documentation, stop dates may be listed in different formats (e.g., mm/dd/yyyy vs. yyyy/mm/dd), racial classifications may be inconsistent (e.g., mixing of “other” and “unknown” designations), and “crime suspected” data is likely to present challenges for interpretation. Nonetheless, with sufficient data manipulation, the data provide detailed information on reported policing activity.

b. NYS DCJS

In addition to data on arrestees and their case outcomes, the Division of Criminal Justice Services compiles information about several other stages of the criminal case process: the probationer population, the jail population at Riker’s Island, and the number of new prison commitments each year from New York City. Data on probationers is available from 1990-2010, the jail population is available from 1995-2010, and the prison commitments data are available from 1994-2010.

c. Land Use Data (multiple sources)

Data on public housing, both in terms of the areal size and location of projects, and the population living in city public housing, is available through the New York City Housing Authority (NYCHA), whose web page provides information on the location and size of public housing projects. Additional information on subsidized housing is available at the State level (regarding Mitchell Lama middle-income projects) and the Federal level (regarding Section 8). Some data on neighborhood composition is also available from NYC.gov’s division of Small Business Services. The “BID Book” contains maps of the city’s Business Improvement Districts.

http://gis.nyc.gov/nycha/im/wmp.do. The underlying database presumably has information in a more easily analyzable form, if we are able to access it.
http://www.dher.state.ny.us/Apps/hsgdevls/hsgdevls.asp
Districts, as well as the dates the BIDs were established. The BID book, publicly available as a .pdf\(^\text{68}\), is current as of 2009. While it does not contain information on BIDs that were established and later eliminated (if any exist), it tracks BID establishment as far back as 1976.

d. NYS Education Department

The New York State Education Department tracks “school and district report cards” from the 1999-2000 school year through the 2009-2010 school year\(^\text{69}\). While these data are only available for the second half of the 1990-2010 crime drop, they may provide some information of interest. The Report Cards provide data for public and charter schools, districts, and New York State as a whole, examining enrollment, demographic, attendance, suspension, dropout, teacher, assessment, accountability, graduation rate, fiscal data, and other information of interest.

3. Commercial and aggregation data sources

Administrative data are kept by government agencies for the purpose of archiving or reporting, but are often not presented in a form conducive to analysis. Several products are commercially or publicly available that aggregate data from other sources, or provide data projections to approximate socioeconomic conditions across the city. One of the most comprehensive subscription-based aggregate sources for the study of New York City and State, described in part above, is Infoshare. In addition to the crime data already discussed, Infoshare contains population data, data on schools, public assistance, hospital admissions, land use, and other social indicators. Depending on the variable of interest, data may be listed by Census Tract, ZIP code, Community District, Neighborhood, or other units.


\(^{69}\) https://www.nystart.gov/publicweb/
The Furman Center, also described above, also collects data on neighborhood conditions, with a focus on factors related to the real estate market (e.g., housing stock, transportation, education). Their Data Search Tool allows the selection of social indicators and the geographic unit and years of primary interest. As with administrative data more broadly, some indicators are tracked as far back as the 1970s, but most are tracked from the 1990s or 2000s forward. Another subscription-based research tool with New York City demographic data is Social Explorer, which compiles Census data over more than 200 years, annual updates from the ACS, data on religion (through the Religious Congregations and Membership Study) and recent (2002) data on carbon emissions. Social Explorer includes a mapping capability, and data are available at a variety of geographic levels, subject to data availability in the original sources.

The New York City Department of Health and Mental Hygiene offers a web-based system called EpiQuery, which permits users to access data from a variety of DOHMH sources. Users can select a single year to view data, or examine trends over time. EpiQuery draws upon several surveys – e.g., the Youth Risk Behavior Survey, the Community Health Survey, the New York City Health and Nutrition Examination Survey (NHANES) – as well as vital statistics data, population estimates, and the Communicable Disease Surveillance System. While the EpiQuery system was developed to provide a user-friendly interface for survey data, researchers may also email the NYC DOHMH for access to the raw data that powers the system, in order to perform more intense calculations that might not be feasible within EpiQuery.

In addition to these data aggregation services, ESRI offers commercially available “Demographic” and “Business Analyst” package that is released annually and contains current-year estimates and five-year projections of population information (e.g., age, race, household

---

70 https://a816-healthpsi.nyc.gov/epiquery/EpiQuery/
71 Stayton, Catherine. 2011. Personal Communication. Email to query@health.nyc.gov.
size and economic status, housing status, etc.). Data are available at the block group level and higher, including both Census units (e.g., tracts, places) and others (e.g., ZIP codes). Projections are based on decennial Census data, and intra-census estimates from the ACS and other sources.

C. Other Data Sources

In addition to the data sources listed above, Appendix C summarizes data sources that contain information of substantive interest, but which are unlikely to be useful for our analysis, as they do not accurately reflect New York City conditions during the 1990-2009 time period.

V. Conclusion

The data sources laid out in this paper are summarized in Appendix D, in a table that details their spatial boundaries and the frequency with which data are available, as well as their substantive contribution. It is notable that with few exceptions such as the vital statistics data, it is rare for data sources to cover the entire time period of interest on an annual basis, and particularly rare for data sources to measure small geographic units on an annual level. Certain trends will likely need to be assessed for shorter portions of the crime decline, data allocated across geographic boundaries, data sources blended over time. However, allocating data across regions, and combining data sources, must be done with extreme caution. Given the dearth of information on systematic similarities and differences between datasets, the decisions on whether and how to reconcile data sources, are by necessity based on researchers’ own assumptions and analytical instincts. While many approaches may be suitable for the analysis of trends, care must be taken to ensure that substantive findings are not sensitive to variations in assumptions about data. Comparisons of cross-sectional patterns across data sources, and of source-specific trends over time, can help to ensure the robustness of analytical findings.
Appendix A. New York City Police Precincts
Appendix B. DCJS Data Request Process for Arrestee Tracking

This appendix summarizes personal communication with Renee Konicki, Data Request Coordinator at the New York State Division of Criminal Justice Services, related to the process of requesting arrestee information for an analysis of arrest outcomes and arrestee trajectories. The DCJS is willing to release criminal history data for use in research projects, but researchers must go through a formal process in order to access the data.

Data Access Procedure

Typically, research teams specify the cohort of individuals for whom they want criminal history record information, including information such as:

- The date range of relevant arrests
- The agency or geographic extent of relevant arrests (e.g., arrests made by the NYPD)
- The types of arrests that are of interest (e.g., felonies and/or misdemeanors).
- Restrictions related to prior criminal histories (e.g., whether interested in all arrestees, or only individuals arrested for the first time).

The feasibility of the data extract will depend on the number of individuals for whom information is requested. For our purposes, the cohort size can be approximated from NYPD arrest data. The research team will also need to provide a legal agreement that specifies the terms of the project, human subjects protections, and a timetable for completion.

With the cohort specified and a legal agreement in place, the DCJS can pull information from either the “top charge” file or the “all charge” file. Renee’s suggestion was to focus on top charges, since the cohort of interest will be large, and the programming requirements for an “all charge” file are likely to outgrow the realm of programming feasibility.

Key variables from the top charge file include:

- Arrest county
- Arresting agency (Not clear from the documentation whether this is down to the precinct level, or just “NYPD” more broadly)
- Arrest date
• Crime date
• Top charge arrest details (charge, weapons, etc.)
• Top charge disposition (and date and agency)
• Arraignment charge and details (upper and lower court)
• Disposition information (court case #, sentence and length)
• Miscellaneous event details (Counsel type, bench warrant, ACD, dismissal)
• Corrections information (admission/release dates, parole/probation dates)
  o From DOCS
  o Can be linked to specific arrest incidents using “Admission reason” field.

Records in the top charge file are identified by “criminal cycle” or arrest incident. The “cycleid” identifier is used internally at DCJS, and may change over time, though will not change within a given file build. Arrestees are identified within each criminal cycle by their New York State ID, or NYSID. However, data can be provided to the research team with the identifying information encrypted to avoid disclosure of personal information. Records also contain a Criminal Justice Tracking Number (CJTN), used to link criminal justice processing events at various stages of an individual criminal cycle (i.e., arrest). Unlike the “cycleid”, which may vary between files, the CJTN is fixed and may also be used by outside criminal justice agencies. However, not all criminal cycles currently have a CJTN assigned to them.

Combining Data Sources

DCJS records have the potential to be merged with other agencies at the NYSID or CJTN level. These merges could happen in one of two ways. Other agencies could send a file identified at the NYSID level, which DCJS could then merge with its own files, encrypting the combined file before providing it to the research team. Alternatively, if the collaborating agencies do not want to release identified information, then the DCJS can provide a crosswalk to encrypted identifiers, and the collaborating agencies could each provide encrypted files, with
data encrypted consistently across agencies. There is precedence for such merges, but cooperation is required from all agencies involved.

**Cost and Timeline**

Cost details are provided in the DCJS Data Request Policy. Fees are determined not by the number of records requested, but by the complexity of programming required to compile the data. A “top charge” file merged with information from other agencies might still be priced according to the top charge file.

The DCJS currently has a backlog of data requests, and between administrative delays and programming time, we would need to allocate at least 2-3 months of turnaround time in order to produce a data file.

**Archiving and Dissemination**

The DCJS requires that projects using their data have “a beginning and end date” and that identified data be destroyed upon the project’s completion. However, they are willing to work with researchers involved in NIJ-funded projects and others that require archiving at NACJD or ICPSR. Archiving and encryption information can be written into the legal agreement that governs the terms of data use.
Appendix C: Data Sources of Limited use for NYC Crime Drop Analysis

This appendix inventories data sources that, while substantively of interest, are unlikely to accurately reflect New York City conditions during the 1990-2009 period.

1. Drug Use Forecasting/Arrestee Drug Abuse Monitoring (DUF/ADAM)

The DUF program was established by the National Institute of Justice in 1988, to validate arrestee self-reports of drug use with an objective drug testing method. Arrestees were also asked about their experiences with guns. 24 cities participated in the program, which was eventually expanded to 35 cities, and continued as the ADAM program in 2000. The ADAM program offered several improvements over the DUF program, including probability sampling, an extended questionnaire, and an expansion to additional sites. Data collection is ongoing, with data available via ICPSR as recently as 2009.

Although the DUF/ADAM data have contained information on New York City since the program’s inception, New York City they are unlikely to adequately represent the drug use and handgun encounters of New York City arrestees. In addition to sampling issues in the early years of the program, more importantly, the catchment area of the New York City DUF/ADAM site is limited to Manhattan. Fewer than half the City’s arrests take place in Manhattan, and differences in policing patterns across the boroughs suggest that Manhattan arrestees are unlikely to be representative of the city as a whole. DUF/ADAM data may therefore be used to examine Manhattan arrestees, and is particularly valuable for the ability to compare Manhattan to other large urban areas across the country. However, the lack of coverage within New York City suggests that the data will be of limited use for a New York City focused study.
2. NYC Criminal Justice Agency (CJA\textsuperscript{72})

The CJA maintains an arrest-level database that includes data on each arrest, case processing, and court outcomes of most New York City arrestees. However, the CJA data are unlikely to be useful for an extended analysis of arrestees. The CJA only tracks court appearances and other key events for as long as the Agency has a need for the information. Data on a given arrest may therefore be incomplete and not include details such as court outcomes or sentencing information. Finally, CJA case records may be removed from the database if they are no longer needed (e.g., in cases of an Adjournment in Contemplation of Dismissal). Data may therefore be incomplete.

3. Behavioral Risk Factor Surveillance System (BRFSS)\textsuperscript{73}

The BRFSS, conducted annually by the CDC since 1984, collects detailed data on respondent behaviors that may affect their health. Questions related to alcohol consumption and “social context” (e.g., neighborhood safety and residential stability) may be of interest for the study of crime. However, it is important to note that particular questions may not be asked in each year; the question on neighborhood safety, for example, is only asked from 1996-1999. Moreover, the BRFSS data are representative of states rather than cities; using prevalence rates from the BRFSS would presume that the City is statistically similar to the rest of New York State. The SMART BRFSS, which began in 2002, contains data at smaller levels of aggregation. Data are available for the “New York – White Plains – Wayne, NY-NJ Metropolitan Division” MSA, and within New York City, data are available for Kings, New York, and Queens Counties, but not Bronx or Richmond Counties. Data in even these sources, however, were collected over

\textsuperscript{72} Freda Solomon at the CJA provided this information in personal communication, 5/20/2011.

the same time period as the New York City Community Health Study, which was designed to provide borough-level estimates. It is

4. Other Administrative Data

Despite the diverse range of data available through city administrative sources, some data of potential interest is only available in recent years, and may not be available for the full period of the crime decline. For example, the New York City Department of Education conducts an annual “school survey” with the potential to provide information on school learning environments and the role schools play in their communities; however, this survey has only been conducted since 2007, and is unlikely to be useful for studying larger trends. Likewise, school “progress reports” and “quality reviews”, contain information on student performance, student progress, and the school environment, but have only been collected since the 2006-2007 school year, and since 2005, respectively.

Data are available online related to “311” calls for nonemergency service. The “311 Online Service Request Map” presents up to a year’s worth of 311 calls for service at varying levels of geography (as large as the city, or as small as a building), and allows searches by location, community board, service request category, and complaint type. However, this data would require a tremendous amount of manual coding, and would require ongoing monitoring to be useful for future analysis. Furthermore, while 311 data may be useful to measure the level of social disorder reported in a given community district, the data were only collected – and 311 service only available – from 2003 forward. Not only is the detailed information on calls for service only available for a single year, data are only archived in citywide aggregates in

74 http://schools.nyc.gov/Accountability/tools/survey/default.htm
75 http://www.nyc.gov/apps/311srmap/
semiannual Mayor’s Management Reports. For these data to be useful in studying calls for service, they would need to be collected on an ongoing basis.