Federal Statistics

- Increased nonresponse to surveys
- Decreased funding
- Demand for more granular data
  - Faster, more frequent
  - More geographic detail
- Demand for more privacy
Crime Statistics

• Brief history
• Quality of current data sources
• Using multiple sources of data
• Measuring accuracy
• Unintended consequences?
Why Crime?

Book for general public:
*Measuring Crime: Behind the Statistics.*

ASA/CRC Press Series on Statistical Reasoning in Science and Society

In print, late 2018
Edith Abbott

• PhD Economics 1905
• Hull House resident
• First female dean of a US graduate school (U Chicago, 1924)
• Statistician, Chicago City Council Committee on Crime, 1914-15

www.ssa.uchicago.edu/edith-abbott
Statistics Relating to Crime in Chicago

by

MISS EDITH ABBOTT, Ph. D., CHICAGO SCHOOL OF CIVICS AND PHILANTHROPY
Abbott (1915)

- Need accurate statistics:
  - Use police resources efficiently
  - Dispel myths about crime
- Vital stats: model for crime stats
- Comparison groups: identify outliers
- Bring together statistics about crime from various sources
The 1920s

CHICAGO – AS THE REST OF THE WORLD MUST IMAGINE IT

Chicago Tribune, 1925
1930 Uniform Crime Reports

- Offenses known to police
- **Uniform** crime definitions
- Originally only counts
Vital Statistics

- 1933: All states
- Homicide only
No one way of describing crime describes it well enough.

Fraud

Survey: capture crimes not in UCR

National Crime Victimization Survey
1990s: CompStat

• Computer-intensive predictive policing
• Data available
  • Quickly
  • Geographic detail
Chicago Violent Crime

Violent crime past 90 days, Feb 18, 2018
Chicago Near West Side, Feb 5 to 18

gis.chicagopolice.org
Fri, 09 Feb 2018 11:29:00

DATE | BLOCK | CODE | CRIME_TYPE | SECONDARY | ARREST | LOCATION | DOMESTIC | BEAT
--- | --- | --- | --- | --- | --- | --- | --- | ---
100 N WESTERN AVE | 031A | ROBBERY | ARMED: HANDGUN | N | GAS STATION | N | 1222
Data Sources: Quality

- Coverage/Response
- Spatial Detail
- Timely
- Transparent

UCR

Police Depts

Vital Stats

NCVS

Homicide
Violent
Property
Fraud
UCR

• Voluntary
• Measurement error
• Only crimes recorded by police
• No measures of uncertainty
Police Depts

- Only large depts.
- Similar to UCR
- More detail
- Faster
- Algorithms secret
Vital Statistics

• Best data we have
• QI program
• National Violent Death Reporting System
• Homicide
Natl Crime Victimization Survey

- Crimes not in other sources
- Small n for victims
NCVS: \[
\frac{\text{Median weight of age group}}{\text{Median weight of age 60+}}
\]
At what response rate is design-based SE invalid?
“A sample that is 95 or 98 percent a probability-sample ... may still be an excellent sample, although it is important to investigate the remaining 5 or even 2 percent as soon as possible.”
NCVS Rape, robbery, agg. assault

Millions of Victimization


Estimate
95% Conf. Int.
NCVS Standard Errors

• Too small
• Only sampling error
• Poststratification
  • lowers SE further
  • assumes bias is removed

• UCR, police stats don’t attempt SE
Common and special causes

• 2006, 2016 anomalies treated as special causes
  • New areas
  • New interviewers

• These are systems features

• Standard errors need to include causes of variation from the system
How?

• Nonresponse adjustments use models
• Include model uncertainty in SE
  • Spiegelhalter & Riesch (2011), risk analysis
  • Lohr & Brick (2017), Bayesian model averaging (1936 *Literary Digest* poll)
• Always larger than design-based SE
How?

• Measurement error models
• Hierarchical models
• Use multiple sources
  • Thurs 10:15: NAS report on methods for combining data
  • Lohr & Raghunathan (2017), *Statistical Science*
Rape, robbery, agg. assault

Millions of Offenses

NCVS Reported to Police

95% Confidence Interval

UCR

1.0

1.5

2.0

2.5

3.0

3.5

<table>
<thead>
<tr>
<th>In UCR</th>
<th>In NCVS</th>
<th>Not in NCVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported to police</td>
<td>Homicide</td>
<td>Reported to police against: Children, institutionalized, commercial establishments</td>
</tr>
<tr>
<td>Except</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not reported to police</td>
<td></td>
<td>Not reported to police: Children, institutionalized, commercial establishments</td>
</tr>
<tr>
<td>Not in UCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not reported to police</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not represented in NCVS:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonresponse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measurement error</td>
<td></td>
</tr>
</tbody>
</table>
Multiple sources

• Statistics from merged data
• Explore error properties
• Present alternative views
• Diversity is a strength
• Alternative sources for lower right quadrant, fraud
• 1960s-70s: expts, checks
Unintended Effects?
Unintended Effects?

• Linking records
  • Need affirmative consent to link survey
  • Public trust, response rates?
• Replacing data with models?
  • What if relationships change?
  • Convenient data sources manipulable
  • Need high-quality surveys or other independent sources
Unintended Effects?

• More honest standard errors?
  • Standards for sensitivity analyses
  • It’s easy to find model giving no nonresponse bias
  • Incentives penalize thorough analysis

• Errors in data sources correlated
  • Victimization missing from all

• Privacy?
Abbott (1915)

• Need accurate statistics:
  • Use police resources efficiently
  • Dispel myths about crime
• Vital stats: model for crime stats
• Comparison groups: identify outliers
• Bring together statistics about crime from various sources
Future Research

• Use multiple sources to study quality
• Standard error:
  • Need systems change
  • Include measurement, nonresponse
• “We of the West are not afraid of crossing the frontiers”