The On-going Challenge of Data Quality

May 2018

JRSA Regional Training Conference

William (Will) Hauser, PhD, Research Analyst - Senior Wisconsin Statistical Analysis Center Bureau of Justice Information and Analysis (BJIA) Wisconsin Department of Justice
What we will be discussing...

1. Quick overview of the Wisconsin SAC
2. The challenge of quality data
3. Initial efforts to improve data quality
4. Current/planned efforts to make it even better
5. Additional thoughts (and challenges)
Wisconsin SAC

- Located within the Wisconsin Department of Justice
- A relatively “new” SAC (in the current structure)

**Major initiatives**
- Uniform Crime Reporting Program
- NCS-X/IBR Transition
- JRSA VOCA-SAC partnership
- Program evaluation for treatment courts and diversion programs
- Sexual Assault Kit Initiative (SAKI)
- Frequent Sobriety Testing Pilot Program
- Domestic Abuse Reporting System
- Criminal history recidivism analysis
- Drug overdose fatality reviews
High Quality Data

- Reliable
- Valid
- Relevant
- Unbiased
- Timely
Challenges to high quality data...

- Administrative records often collected for another purpose
  - Not intended to be aggregated as ‘data’
  - Failure to understand the potential value or utility of the records treated as ‘data’

- Often reliant on multiple external entities and individuals to enter or provide data
  - Requires agency commitment to quality data
  - Often leads to manual follow-up
  - Difficult to control what you receive
  - No easy way to communicate the issues and resolve them
  - Staff turnover and competing obligations

- Data is increasingly public
  - Relevance
  - Correct interpretation
  - Timeliness
Common problems

Identifying them is not always easy...get to know your data

- Lack of uniform collection or reporting methods
- Missing data and data black holes
- Illogical and implausible values
- Incorrect classification and/or lack of uniform classification scheme
- Intrinsic limits
### Offenses

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggravated Assault</td>
<td>21</td>
<td>16</td>
<td>12</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>Simple Assault</td>
<td>127</td>
<td>122</td>
<td>121</td>
<td>146</td>
<td>127</td>
</tr>
</tbody>
</table>

### Arrests

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggravated Assault</td>
<td>171</td>
<td>179</td>
<td>162</td>
<td>195</td>
<td>184</td>
</tr>
<tr>
<td>Simple Assault</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>
Real data from a WI agency

<table>
<thead>
<tr>
<th>Offense</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorderly Conduct</td>
<td>88</td>
<td>121</td>
<td>8</td>
</tr>
<tr>
<td>Driving Under the Influence</td>
<td>652</td>
<td>874</td>
<td>6</td>
</tr>
<tr>
<td>Grand Total</td>
<td>1010</td>
<td>1349</td>
<td>435</td>
</tr>
</tbody>
</table>
Efforts to improve data quality

- **Develop automated data quality reports**
  - Annually sent to each agency/county/site
  - Flag missing data, logical inconsistencies, and implausible values or trends
  - Follow-up to ensure that report is viewed – enforce accountability

- **Fixing identified problems is not enough**
  - Create new data validations
  - Improve existing validations by testing for holes
  - Use problem cases to improve or correct data validations
Efforts to improve data quality

- Manually review individual records against aggregate data to identify systemic issues
- Make the data more readily available, visible, and digestible
- Sometimes you have to work with what you have
  - Document!
- Training...
- Visualizations to explain data collection process
What kind of sexual offense is it?

Did everyone who participated in the sexual act consent to the act(s)?

Consider age (cannot consent under age 10) and impairment due to physical or mental disability or intoxication.

1. Was anyone under age 18?
   yes -> Statutory Rape
   no -> 2. Were the participants blood relatives?

   yes -> Incest
   no -> 1. Did penile-vaginal penetration occur or was it attempted?

       no -> 2. Did oral or anal sex occur or was either attempted?

           no -> Sodomy/Oral Sex
           yes -> 3. Did non-penile penetration of the anus or vagina occur or were either attempted?

               no -> Assault with Object
               yes -> 4. Did the aggressor touch the victim's private parts for sexual gratification?

                   yes -> Fondling
                   no -> Assault with Object

Report two cases if both participants were under 18 once each as victim and as offender.

To a degree within which marriage is prohibited by law: siblings, first cousins, etc. Cannot be nearer in kin than 2nd cousins.

If at least one of the participants in the sexual act(s) was under 18 and they were closely related, report as both statutory rape and incest.

Includes intercourse as well as anal/oral sex and insertion of an object.

Consider reporting as simple assault (138) if a non-sexual form of unwanted contact occurred or the fondling was for reasons other than sexual gratification (e.g., to embarrass).
What level of assault is it?

- Was a dangerous weapon such as a gun or knife threatened OR used? yes → Aggravated Assault
  no → Was the victim seriously injured? yes → Simple Assault
    no → Was the victim at risk of being seriously injured, or did the offender intend to seriously injure the victim? yes → Disorderly Conduct
      no → Did the subject offend, shock, or behave inappropriately?

- Was the victim seriously injured? yes → Simple Assault
  no → Consider: Would a reasonable person have sought medical attention? Was the victim’s breathing restricted?

- Were hands/feet/teeth or another non-dangerous weapon used on the victim? yes → WIBRS: Intimidation
  no → Consider: coercion, hazing, stalking, bomb threat

- Was the victim threatened and in fear of bodily harm? yes → WIBRS: Intimidation
  no → Consider: coercion, hazing, stalking, bomb threat
Current/planned efforts

- Automated data quality reports
  - Building into web applications

- New data collection efforts
  - With lots of validations
  - Has its challenges too

- Work directly with the solution provider and the agency

- Verification across multiple sources
  - Compare criminal history to other data sources

- Recognize it is an on-going process!
Questions?

Will Hauser, PhD
Research Analyst - Senior
608-261-7449
hauserwj@doj.state.wi.us