# THIRD NATIONAL JUVENILE ONLINE VICTIMIZATION INCIDENCE STUDY (NJOV-3)

## NDACAN Dataset Number 209
## USER'S GUIDE



**National Data Archive on Child Abuse and Neglect**
**Bronfenbrenner Center for Translational Research**
**Cornell University**
**Ithaca, New York 14853-4401**
**607-255-7799**
**ndacan@cornell.edu**

**https://www.ndacan.cornell.edu**

Initial release 02/16/2018

# THIRD NATIONAL JUVENILE ONLINE VICTIMIZATION INCIDENCE STUDY (NJOV-3)

## Data Collected by

Janis Wolak, J.D.
University of New Hampshire Durham, NH

David Finkelhor, Ph.D.
University of New Hampshire Durham, NH

Kimberly J. Mitchell, Ph.D.
University of New Hampshire Durham, NH

## Funded by

## Distributed by

National Data Archive on Child Abuse and Neglect

## User's Guide and Codebook Written by

Holly M. Larrabee, MSHP
National Data Archive on Child Abuse and Neglect
in collaboration with
Janis Wolak, J.D.
University of New Hampshire Durham, NH

# Table of Contents

# Preface

The data for *Third National Juvenile Online Victimization Incidence Study (NJOV-3)*, have been given to the National Data Archive on Child Abuse and Neglect for public distribution by Janis Wolak. Funding for the project was provided by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice (Award Number: 2009-SNB-90001).

# Acknowledgement of Source

Authors should acknowledge the National Data Archive on Child Abuse and Neglect and the original collector of the data when they publish manuscripts that use data provided by the Archive. Users of these data are urged to follow some adaptation of the statement below.

> The data used in this publication were made available by the National Data Archive on Child Abuse and Neglect, Cornell University, Ithaca, NY, and have been used with permission. Data from *Third National Juvenile Online Victimization Incidence Study (NJOV-3)* were originally collected by Janis Wolak, David Finkelhor, and Kimberly J. Mitchell.  Funding for the project was provided by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice (Award Number: 2009-SNB-90001). The collector of the original data, the funder, NDACAN, Cornell University and their agents or employees bear no responsibility for the analyses or interpretations presented here.

The bibliographic citation for this data collection is:

> Wolak, J., Finkelhor, D, & Mitchell, K. J. (2018). *Third National Juvenile Online Victimization Incidence Study (NJOV-3)* [Dataset]. Available from National Data Archive on Child Abuse and Neglect Web site, https://www.ndacan.cornell.edu/.

## Publication Submission Requirement

In accordance with the terms of the *Data License* for this dataset, users of these data are required to deposit a copy of any published work or report based wholly or in part on these data with the Archive. A copy of any completed manuscript, thesis abstract, or reprint should be sent to the National Data Archive on Child Abuse and Neglect, Cornell University, Bronfenbrenner Center for Translational Research, Beebe Hall, Ithaca, New York 14853. Such copies will be used to provide funding agencies with essential information about the use of NDACAN resources and to facilitate the exchange of information about research activities among data users and contributors.

# ABSTRACT

The 3rd National Online Victimization Study (NJOV-3) is the third wave of a longitudinal study. Wave 1 (NJOV-1) pertained to arrests for technology-facilitated crimes (e.g., sex offenders using the Internet to meet minors, solicitations to undercover investigators posing online as minors, downloading of child pornography) that occurred between July 1, 2000 and June 30, 2001; Wave 2 (NJOV-2) pertained to arrests during 2006. NJOV-3 collected data about technology-facilitated child sexual exploitation crimes ending in arrest in 2009 .

For each wave of the NJOV Arrest Study, data was collected in two phases. Phase 1 was mail survey of a national sample of law enforcement agencies that asked if agencies had made arrests for technology-facilitated child sexual exploitation crimes during a specific time frame.  Phase 2 was telephone interviews with investigators to collect details about individual cases reported in the mail surveys. The goal of this methodology was to 1) utilize a representative national sample of law enforcement agencies that would give us an overall picture of these crimes in the United States, 2) understand how these cases emerged and were handled in a diverse group of agencies, 3) get detailed data about the characteristics of these crimes from well-informed, reliable sources, and 4) see how the prevalence and characteristics of such crimes may have changed over time.

# STUDY OVERVIEW

## Study Identification

*Third National Juvenile Online Victimization Incidence Study (NJOV-3)*

### Investigator(s):

Janis Wolak, J.D., University of New Hampshire Durham, NH

David Finkelhor, Ph.D., University of New Hampshire Durham, NH

Kimberly J. Mitchell, Ph.D., University of New Hampshire Durham, NH

### Funding Agencies:

Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice

### Award Numbers: 2009-SNB-90001

## Purpose of the Study

Research objectives for the study:

- Assess the scope and characteristics of trends in technology-facilitated CSE crimes.
- Examine perpetrator characteristics to illuminate links between possession or trafficking of CP and actual abuse of a child.
- Gather and analyze data about how different technologies are used in CSE crimes and about investigative strategies and techniques that address these issues.
- Examine the characteristics of and strategies for handling cases in which juveniles manufacture or distribute sexually explicit images of themselves or peers.
- Collect data about and evaluate prosecution strategies in technology-facilitated CSE crimes.

## Study Design

Data were collected in two phases. Phase 1 was mail survey of a national sample of law enforcement agencies that asked if agencies had made arrests for technology-facilitated child sexual exploitation crimes during 2009. Phase 2 was telephone interviews with investigators to collect details about individual cases reported in the mail surveys. The goal of this methodology was to 1) utilize a representative national sample of law enforcement agencies that would give us an overall picture of arrests for these crimes in the United States, 2) understand how these cases emerged and were handled in a diverse group of agencies, 3) get detailed data about the characteristics of these crimes from well-informed, reliable sources, and 4) see how the prevalence and characteristics of such crimes may have changed over time.

NJOV-3 was conducted with the approval of the University of New Hampshire's Institutional Review Board and Privacy Certificates were filed in accordance with regulations mandated for research funded

by the U.S. Department of Justice.

## Date(s) of Data Collection

Data collection started in January 2010 and concluded in March 2011 regarding arrests that happened during 2009.

## Geographic Area

United States

## Unit of Observation

Arrests for technology-facilitated sex crimes committed against children.

## Sample

Mail Survey

The NJOV3 sample of law enforcement agencies was designed to yield a nationally representative sample of technology-facilitated child sexual exploitation cases (i.e., sex offenders using the Internet or other electronic technologies to facilitate sex crimes with juvenile victims, solicitations to undercover investigators posing online as minors, child pornography possession and distribution) that ended in arrest. We used a three-frame stratified sample of agencies because arrests for such crimes do not occur with equal probability among the more than 15,000 U.S. law enforcement agencies. We divided agencies into three sampling frames based on their specialization, training or known experience with technology-facilitated sex crimes against minors. This was done so that we could get information from the agencies that were most likely to see these types of crimes while still allowing every agency a chance to be selected for the sample. The sample of agencies was recruited and classified into frames for the first NJOV Study, with some changes made in the second and third wave to accommodate alterations in the status of sampled agencies (e.g., some agencies began to specialize in technology-facilitated crimes; some ceased to exist; some were added to the first frame because they were known to have experience investigating certain types of cases).

The NOV3 first frame sample consisted of 176 agencies. These included 61 Internet Crimes Against Children (ICAC) Task Forces, which were funded by the US Department of Justice to investigate technology-facilitated child sexual exploitation crimes; 38 agencies which were satellite ICAC Task Forces during Wave 1; 16 agencies which were known to work with civilian groups that conducted independent investigations of technology-facilitated crimes (included for the NJOV-2 Study), and 61 agencies that, according to media reports, had handled specific types of cases relevant to the study.

 We did not sample first frame agencies; rather, we surveyed all of them. In addition, two federal agencies were included in the study "with certainty." These were not counted in the mail survey sample because they reported their cases electronically rather than through mail surveys due to large caseloads. Of the 176 first frame agencies that received mail surveys, 3% (n=5) lacked jurisdiction. Of the 171 eligible agencies, 81% (n=138) returned completed surveys and 57% (n=97) reported making one or more arrests for technology-facilitated child exploitation crimes.

The second frame consisted of agencies that we considered more likely than other agencies to have investigated technology-facilitated child sexual exploitation crimes because they had staff that had received training in these types of cases. We identified these trained agencies prior to NJOV-1 by using lists of agencies participating in week-long training programs. The lists were acquired from two training organizations, SEARCH and the National Center for Missing & Exploited Children. Also, an additional agency from a large metropolitan area was included with certainty in the second frame to ensure that agencies from all major metropolitan areas in the U.S. were included in the sample. There were 1,636 agencies in the NJOV-3 Study second frame population, with approximately 50% selected to receive the mail survey. (These agencies were selected in Wave 1 (NJOV-1) and resurveyed in NJOV-2 and NJOV-3.)

Of the second frame agencies that received mail surveys in Wave 3 (n=815), 3% (n=28) were ineligible to participate because they lacked jurisdiction to investigate technology-facilitated child sexual exploitation crimes or the agency no longer existed. (In this category were mostly small town agencies that relied on county or other larger jurisdictions to conduct criminal investigations.) Of the 787 eligible second frame agencies, 84% (n=659) completed and returned mail surveys, and 35% (n=275) reported making one or more arrests for technology-facilitated child exploitation crimes.

The third frame consisted of 13,572 other local, county and state law enforcement agencies across the US. The third frame sample was drawn using a database available through the National Directory of Criminal Justice Data (National Directory of Law Enforcement Administrators, 2009). This data set included an annually updated census of local, county, and state law enforcement agencies and was designed to provide geographic and other identifying information for each record included in either the FBI's Uniform Crime Reports files or the Bureau of Justice Statistic's Directory of Law Enforcement Agencies. The agencies in the first and second frames were cross-referenced with third frame agencies in the database to avoid duplication among the three frames. Of the 13,572 third frame agencies, 12% (n=1,662) were randomly selected to participate in the study. (These agencies were selected in NJOV-1 and resurveyed in NJOV-2 and NJOV-3.) Of the 1,662 third frame agencies that received mail surveys in Wave 3, 9% (n=157) were ineligible to participate because they lacked jurisdiction, the agency no longer existed, or the agency was a duplicate of another agency within the third frame. Of the 1,505 eligible agencies, 88% (n=1,331) completed and returned mail surveys, and 14% (n=218) reported one or more arrests for technology-facilitated child sexual exploitation crimes.

Telephone interviews

We conducted detailed telephone interviews with law enforcement investigators to gather information about arrests for technology-facilitated crimes that were reported in the mail survey.  The 2,128 agencies that responded to the NJOV3 mail survey, plus the two federal agencies that participated reported 4,010 arrests for technology-facilitated child sexual exploitation crimes in 2009, with 97 first frame agencies reporting 2,342 arrests, 275 second frame agencies reporting 1,141 arrests and 218 third frame agencies reporting 527 arrests.

We designed a sampling procedure that took into account the number of cases reported by an agency, so that we would not unduly burden respondent investigators in agencies with many cases.  If an agency reported between one and three arrests, we conducted follow-up interviews for every case.  Sixty-one percent of the agencies that made arrests in relevant cases in NJOV-3 fell within this group, compared to 85% in Wave 1 and 71% in Wave 2. For agencies that reported more than three cases, we conducted

interviews for all cases that involved identified victims (victims who were located and contacted during the investigation), and sampled other cases. For agencies with between four and fifteen cases, half of the cases that did not have identified victims were randomly selected for follow-up interviews. In agencies that reported more than fifteen cases, cases with no identified victims were divided into two samples, using random selection, and then half of one sample was randomly selected for follow-up interviews. In some agencies, we could not find out which cases had identified victims before we selected cases for the sample, so we sampled from all cases, using the procedure described above.

Of the 4,010 cases reported in all three frames of NJOV-3, 38% (n=1,522) were not selected for the telephone interview sample and 11% (n=459) did not meet eligibility requirements for the study (i.e., arrests in 2009 for technology-facilitated sexual exploitation crimes with victims age 17 or younger, including unidentified victims such as those pictured in child pornography or victims portrayed by undercover investigators posing online as minors.) There were 2,029 eligible cases in the sample. The overall response rate for telephone interviews was 64%.

## Data Collection Procedures

Procedures for the Phase 1 mail survey of law enforcement agencies

To maximize response rates to the Phase 1 mail survey, we followed an adapted version of the "total design" mail survey methodology (Dillman, 2007). These procedures were used in all three waves of the NJOV Study.

1. We sent surveys, personalized cover letters, and business reply envelopes to the heads of the local, county, and state law enforcement agencies in the sample.

2. Approximately three weeks after the initial mailing, we sent reminder postcards to all agency heads, asking them to complete and return the survey if they had not done so, and thanking them if they had.

3. Approximately six weeks after the initial mailing, we sent additional copies of the survey, personalized cover letters, and business reply envelopes to the heads of agencies who had not responded to date.

4. Approximately nine weeks after the initial mailing, we sent third copies of the survey, personalized cover letters, and business reply envelopes to agency heads that had still not responded.

5. Finally, telephone interviewers called the agencies that had not responded and, when possible, filled out the survey instrument over the telephone.

The mail survey began in January 2010. Mailings concluded in April 2010; calls to non-responders continued while telephone interviews were being conducted - until March 2011.

Procedures for the Phase 2 telephone interviews about specific cases

Seven trained interviewers conducted the NJOV-3 telephone interviews, using a computer-assisted telephone interviewing system, WINCATI. The interview was programmed to include: 1) question and response series; 2) skip patterns; 3) interviewer probes and instructions; 4) range checks; and 5) special edit procedures. The WINCATI system provides a number of benefits over traditional telephone

interviewing, including a smoother flowing interview, the ability to branch to different series of questions depending on answers to previous questions, and automated skip patterns.

The interviewers attended a two-day training session led by the researchers that provided extensive details about the background, purpose, and instrumentation of the study, and they participated in a series of observational and practice interviews. The telephone interviews for the NJOV-3 Study were conducted between March 2010 and March 2011.

## Response Rates

The overall response rate for the NJOV-3 mail survey was 86%. The response rate for telephone interviews with law enforcement investigators about specific cases was 64%.

## Sources of Information

Telephone Interview. No data were generated from the mail survey, as its sole purpose was to determine if agencies had eligible cases.

## Type of Data Collected

Survey.

## Measures

Copies of the measures are included in the dataset materials and are not to be used in a primary data collection without expressed written permission from the study authors.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2010). *Third National Juvenile Online Victimization Mail Survey*. Durham, NH: University of New Hampshire.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2010). *Third National Juvenile Online Victimization Telephone Survey*. Durham, NH: University of New Hampshire.

## Related Publications & Reports

**Users are strongly encouraged to review these references before doing analyses.  To view a complete list of publications visit our online citations collection called "canDL" by going to** https://www.ndacan.cornell.edu/publications/publications.cfm **, Once on the page, navigate to the DS# 209 folder to view all publication citations relevant to this dataset.**

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2011). *Methodology Report: 3rd National Juvenile Online Victimization (NJOV3) Study*. Durham, NH: Crimes against Children Research Center, University of New Hampshire.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012). *Trends in Law Enforcement Responses to Technology-facilitated Child Sexual Exploitation Crimes: The Third National Juvenile Online Victimization Study (NJOV-3)*. Durham, NH: Crimes against Children Research Center.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012). *Trends in Arrests for Child Pornography Possession: The Third National Juvenile Online Victimization Study (NJOV-3)*. Durham, NH: Crimes against Children Research Center, University of New Hampshire.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2012). *Trends in Arrests for Child Pornography Production: The Third National Juvenile Online Victimization Study (NJOV-3)*. Durham, NH: Crimes against Children Research Center.

Wolak, J., & Finkelhor, D. (2013). Are crimes by online predators different from crimes by sex offenders who know youth in-person? *Journal of Adolescent Health, 53*(6), 736-741. doi: 10.1016/j.jadohealth.2013.06.010

Wolak, J., & Finkelhor, D. (2013). *Trends in Arrests for Technology-Facilitated Sex Crimes with Identified Victims: The Third National Juvenile Online Victimization Study (NJOV-3)*. Durham, NH: Crimes against Children Research Center.

## Analytic Considerations

Missing data is coded as such on a variable by variable basis within the value labels field. Some missing data is coded as system missing.

### Weighting

Because of the sampling design of this study, the dataset MUST be analyzed using complex samples survey procedures, which are available in most major statistical analysis software. Analyzing the data without accounting for the cluster design will lead to inaccurate results. The weight variable is needed to derive appropriate point statistics (means, frequencies, etc.) and the stratum and psu variables are need to compute proper standard errors.

The weighting syntax for STATA:

svyset psu [pw=pweight], singleunit(cen) strata(ltstrata) fpc(fpc1) || ssu, strata(ltstrata2) fpc(fpc2)

The Complex Samples Analysis set-up for SPSS, a csaplan file has also been included in the dataset:

```
CSPLAN ANALYSIS
/PLAN FILE=
'C:\temp\DS209-Analysis-Plan.csaplan'
/PLANVARS ANALYSISWEIGHT=pweight
 /PRINT PLAN
  /DESIGN STRATA=ltstrata CLUSTER=psu
 /ESTIMATOR TYPE=EQUAL_WOR
/INCLPROB VARIABLE=fpc1
/DESIGN STRATA=ltstrata2 CLUSTER=ssu
/ESTIMATOR TYPE=EQUAL_WOR
 /INCLPROB VARIABLE=fpc2.
```

Below is an example of a Proc Surveymeans set-up for SAS, where the data file referenced in the

"RATE=" statement is a data file that contains the "ltstrata" (strata from the first stage of sampling) and finite population correction varaible (fpc1) which was renamed "_rate_". The SAS documentation for the Proc Survey procedures only requires the first stage of the weighting design to be specified in a multi-stage design dataset. See the official SAS documentation for more information.

```
Proc Surveymeans data=library.figurethree RATE=library._RATE_ mean std clm ;
cluster psu ;
strata ltstrata;
weight pweight;
variable onlinecpp_x3;
run;
```

Combing the NJOV-1, NJOV-2, and NJOV-3 datasets would require a recalculation of the weights and therefore is not recommended.

**A note about replicating summary statistics from published reports:** It may not be possible to replicate all summary statistics found in the published reports produced by the study investigators. The study investigators collected more variables than what was provided in this dataset. This means that variables, upon which the summary statistics were based, may not be available in the archived data. In order to help secondary analysts establish a level of comfort with the application of the weights, we are providing syntax to replicate Figure 1 from the report titled, "Trends in Arrests for Child Pornography Production: The Third National Juvenile Online Victimization Study (NJOV-3)" published in April 2012. You will note that a recode is conducted at the start of the syntax. This recode was brought about by a review of the case narrative (not included in this dataset) conducted by the original researchers. According to correspondence with study staff, "…The reasons for this recode could have been due to: being incorrect based on the case narratives. Some of the recodes involved cases where offenders took suggestive images of children which were not child pornography in the arresting jurisdictions -- for example, they took pictures of kids in public places like gymnastic events or on playgrounds. Some of these cases may have had identified victims if the offender also sexually abused a child, but others may not have had such victims. Other cases involved multiple offenders or victims in which the data should have been consistent with the role of the primary offender or what happened to the primary victim. For example, one offender may have produced child pornography but the primary offender did not, or one victim may have been photographed but the primary victim was not." (personal communication, January 13th, 2018)

As an example of how the cpprod_3 recodes impact the weighted counts, when the recodes are not done, the "All cp production arrests" weighted count is 1,998. The "youth-produced images" weighted count is not impacted by the presence or absence of recoding and will match the report (1,198). The "adult-produced images" weighted count is 800.8. In summary, the recoding of the cpprod_3 for those twelve cases results in a change in the two of the three weighted counts by 88.

**STATA** syntax to reproduce the weighted counts for Figure 1 for year 2009:

*recode values for specific cases.*
recode cpprod_3 1=0 if (caseID==1005.013 | caseID==1089.002 | caseID==1122.005 |
    caseID==1124.014 | caseID==1124.050 | caseID==1132.006 | caseID==1135.013 |
    caseID==1151.015 | caseID==2706.001 | caseID==3219.002 | caseID==3779.001)
recode cpprod_3 1=0 if caseID==1084.151

*Number of arrests "all cp production arrests" in report.*
svy linearize : tabulate cpprod_3, count obs ci

**Results:**
. svy linearize : tabulate cpprod_3, count obs ci se
(running tabulate on estimation sample)

Number of strata  =     3        Number of obs    =    1,299
Number of PSUs   =    477       Population size  = 8,144.4055
                                Design df      =      474

| Case Included CP Production (pre9b) | count | se | lb | ub | obs |
|---|---|---|---|---|---|
| No | 6234 | 359.9 | 5527 | 6941 | 912 |
| Yes | 1910 | 105 | 1704 | 2117 | 387 |
| Total | 8144 | | | | 1299 |

Key:  count   = weighted count
   se     = linearized standard error of weighted count
   lb     = lower 95% confidence bound for weighted count
   ub     = upper 95% confidence bound for weighted count
   obs    = number of observations

*Number of arrests, "youth produced images" (ypi) cases*
svy linearized : tabulate cpprod_3 if ypi_3==1, count obs ci

**Results:**
   . svy linearized : tabulate cpprod_3, subpop (ypi_3) count obs ci se
   (running tabulate on estimation sample)

Number of strata  =     3        Number of obs    =    1,299
Number of PSUs   =    477       Population size  = 8,144.4055
                                Subpop. no. obs  =      228
                                Subpop. size    = 1,197.6347
                                Design df      =      474

| Case included CP production (pre9b) | count | se | lb | ub | obs |
|---|---|---|---|---|---|
| No | 0 | 0 | 0 | 0 | 0 |
| Yes | 1198 | 85.54 | 1030 | 1366 | 288 |
| Total | 1198 | | | | 288 |

Key:  count    = weighted count
 se      = linearized standard error of weighted count
 lb      = lower 95% confidence bound for weighted count
 ub      = upper 95% confidence bound for weighted count
 obs     = number of observations

*Number of arrests for CP production, "adult produced images"-first create a dichotomous variable named ypi_3adult based off from ypi_3, where ypi_3adult =1 means that it was not a youth produced image and therefore was an adult produced image..*
gen ypi_3adult=0
    replace ypi_3adult=1 if ypi_3==0
svy linearized : tabulate cpprod_3, subpop (ypi_3adult) count obs ci se

**Results:**
    svy linearized : tabulate cpprod_3, subpop (ypi_3adult) count obs ci se
    (running tabulate on estimation sample)

Number of strata  =     3          Number of obs    =    1,299
Number of PSUs    =    477          Population size   = 8,144.4055
                                    Subpop. no. obs   =    1,071
                                    Subpop. size     = 6,946.7709
                                    Design df        =    474

| Case included CP production (pre9b) | count | se | lb | ub | obs |
|---|---|---|---|---|---|
| No | 6234 | 359.9 | 5527 | 6941 | 912 |
| Yes | 712.6 | 68.05 | 578.9 | 846.3 | 159 |
| Total | 6947 | | | | 1071 |

**SPSS** syntax to reproduce the weighted counts for Figure 1 for year 2009:

    Do If (caseID=1005.013 | caseID=1089.002 | caseID=1122.005 | caseID=1124.014 |
    caseID=1124.050 | caseID=1132.006 | caseID=1135.013 | caseID=1151.015 | caseID=2706.001 |
    caseID=3219.002 | caseID=3779.001).
    Compute cpprod_3=0.
    end if.
    execute.

    do if (caseID=1084.151).
    recode cpprod_3 (1=0).
    end if.
execute.

*All CP production arrests.*
    CSTABULATE
     /PLAN FILE=
  'C:\temp\DS209-Analysis-Plan.csaplan'
     /TABLES VARIABLES=cpprod_3
     /CELLS POPSIZE TABLEPCT
     /STATISTICS SE CIN(95) COUNT
 /MISSING SCOPE=TABLE CLASSMISSING=INCLUDE.

**Results:**

**Case included CP production (pre9b)**

| | | Estimate | Standard Error | 95% Confidence Interval | | Unweighted Count |
| | | | | Lower | Upper | |
|---|---|---|---|---|---|---|
| Population Size | No | 6234.150 | 359.728 | 5527.291 | 6941.008 | 912 |
| | Yes | 1910.256 | 104.637 | 1704.645 | 2115.866 | 387 |
| | Total | 8144.406 | 358.640 | 7439.685 | 8849.127 | 1299 |
| % of Total | No | 76.5% | 1.5% | 73.4% | 79.4% | 912 |
| | Yes | 23.5% | 1.5% | 20.6% | 26.6% | 387 |
| | Total | 100.0% | 0.0% | 100.0% | 100.0% | 1299 |

*CP production arrests-Youth-produced images and Adult-produced images..*
CSTABULATE
    /PLAN FILE=
  'C:\temp\DS209-Analysis-Plan.csaplan'
   /TABLES VARIABLES=cpprod_3
   /SUBPOP TABLE=ypi_3 DISPLAY=LAYERED
   /CELLS POPSIZE TABLEPCT

/STATISTICS SE CIN(95) COUNT
/MISSING SCOPE=TABLE CLASSMISSING=INCLUDE.

**Results:**
Subpopulation Tables

**Case included CP production (pre9b)**

| Case involved youth produced sexual images (pre6, pre9b1) | | | Estimate | Standard Error | 95% Confidence Interval | | Unweighted Count |
|---|---|---|---|---|---|---|---|
| | | | | | Lower | Upper | |
| No | Population Size | No | 6234.150 | 359.728 | 5527.291 | 6941.008 | 912 |
| | | Yes | 712.621 | 67.803 | 579.389 | 845.853 | 159 |
| | | Total | 6946.771 | 359.261 | 6240.829 | 7652.712 | 1071 |
| | % of Total | No | 89.7% | 1.1% | 87.4% | 91.7% | 912 |
| | | Yes | 10.3% | 1.1% | 8.3% | 12.6% | 159 |
| | | Total | 100.0% | 0.0% | 100.0% | 100.0% | 1071 |
| Yes | Population Size | Yes | 1197.635 | 85.239 | 1030.141 | 1365.128 | 228 |
| | | Total | 1197.635 | 85.239 | 1030.141 | 1365.128 | 228 |
| | % of Total | Yes | 100.0% | 0.0% | 100.0% | 100.0% | 228 |
| | | Total | 100.0% | 0.0% | 100.0% | 100.0% | 228 |

**SAS** code to reproduce the weighted counts for Figure 1 for year 2009:

```
libname library 'C:\temp';
run;
OPTIONS nofmterr;

Data library.DS209_recode;
   set library.DS209;
      if (caseID=1005.013) or (caseID=1089.002) or (caseID=1122.005) or
(caseID=1124.014) or (caseID=1124.050) or (caseID=1132.006)
or (caseID=1135.013) or (caseID=1151.015) or (caseID=2706.001) or (caseID=3219.002)
or (caseID=3779.001)THEN cpprod_3=0;
      if (caseID=1084.151) THEN cpprod_3=0;
run;

Data library._Rate_ ;
      set library.DS209  (keep = caseid fpc1 ltstrata ltstrata2 rename=
fpc1=_RATE_);
      run;

*creates weighted count for "Total cp production" "youth produced images" and
"adult produced images" arrests;

Proc Surveyfreq data=library.ds209 RATE=library._RATE_ ;
cluster psu ;
strata ltstrata;
weight pweight;
table cpprod_3 * ypi_3 / COL CL CLWT;
```

```
run;
```

**Results:**

*for this example, output from SAS has been modified so that it would fit onto the page.

| Data Summary | |
|---|---|
| Number of Strata | 3 |
| Number of Clusters | 477 |
| Number of Observations | 1299 |
| Sum of Weights | 8144.41 |

**Table of cpprod_3 by ypi_3**

| cpprod_3 | ypi_3 | Frequency | Weighted frequency | Std Dev of wgt freq | 95% Confidence Limits for Wgt Freq | | Percent | Std Err of percent |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 912 | 6234 | 355.46336 | 5536 | 6933 | 76.5452 | 1.4694 |
| | 1 | 0 | . | . | . | . | . | . |
| | Total | 912 | 6234 | 355.46336 | 5536 | 6933 | 76.5452 | 1.4694 |
| 1 | 0 | 159 | 712.62096 | 63.17218 | 588.48879 | 836.75312 | 8.7498 | 0.8373 |
| | 1 | 228 | 1198 | 81.89924 | 1037 | 1359 | 14.705 | 1.1259 |
| | Total | 387 | 1910 | 99.20201 | 1715 | 2105 | 23.4548 | 1.4694 |
| Total | 0 | 1071 | 6947 | 356.1224 | 6247 | 7647 | 85.295 | 1.1259 |
| | 1 | 228 | 1198 | 81.89924 | 1037 | 1359 | 14.705 | 1.1259 |
| | Total | 1299 | 8144 | 356.72998 | 7443 | 8845 | 100 | |

Sexting cases involving children were collected as a part of the study, however, they were not archived with this dataset.

**Subsetting:** When analyzing data from a complex survey design, secondary analysts should always use the subpopulation commands for each statistical software program instead of using commands or techniques (manually deleting/dropping observations) designed to remove records from the analyses. Below, is a basic table of commands for each of the programs and their more appropriate counterparts

for the purposes of preserving observations in a weighted analyses. For this dataset, your total weighted population should always be 8,144. To learn more about techniques for examining data from complex sampling designs, please reach out to your institution's statistical consulting department.

| Statistics Software Package | Sub-commands that drop/remove observations from the analyses | Preferred sub-commands for weighted analyses |
|---|---|---|
| STATA | if | subpop |
| SPSS | if | subpop |
| SAS | by | domain |

## Confidentiality Protection

The dataset underwent a confidentiality review by NDACAN and it was determined that no recodes were necessary. There are no primary or seconday identifiers in the dataset.

## Extent of Collection

This collection consists of the User's Guide, Methodology Report, telephone interview survey instrument, Data Dictionary, complex samples analysis plan file for SPSS, one text data file (DS209) with import program files for SAS, SPSS, and Stata, one tab-delimited data file, and files native to SPSS, Stata, and SAS.

## Extent of Processing

NDACAN produced the User's Guide, Codebook, the SPSS, Stata, and SAS program import files, and text data file.

The data contained in this dataset represent the responses received from law enforcement agencies and have been largely untouched, other than confidentiality recodes,  performed by the data contributor.  In addition to the data collected during the telephone survey, study staff collected case narratives. The case narratives allowed study staff to make changes to the data used to produce published articles and reports. Any changes made by study staff have NOT been included in this dataset. NDACAN does not have access to the case narratives due to the sensitive idenfying information contained within them. The implications for this, is that many of the figures and summary statistics presented in the published works, produced by the data contributor, will not be able to be replicated using the archived data.

# DATA FILE INFORMATION

## File Specifications

There is only one data file and it is called DS209, followed by a "v" and a number indicating the current version of the data. It contains the data collected during the telephone survey about arrests that happened during 2009. There are 608 variables and 1,299 records in the data file.

## Data File Notes

There are variables which do not have value labels assigned. The value of "0" represents "NO" and the value of "1" represents "YES."

Only the data from the Arrest study are included in this dataset. The data for the Prosecutor and Sexting portions of this study were not archived.

The term "identified victim" means that a victim was located and contacted as part of the investigation. Minors depicted in child pornography are also victims, but many remain unidentified; hence the differentiation between identified and unidentified victims.

*Table 1. Acronyms and abbreviations used in the dataset*

| Acronym or Abbreviation | Definition or Meaning |
|---|---|
| V | Victim |
| ICAC | Internet Crimes Against Children Task Force Program |
| agy | Agency |
| CSEC | Commercial Sexual Exploitation of Children |
| LEA | Law Enforcement Agency |
| SEM | Sexually Exploited Minors |
| UC | undercover |

| Acronym or Abbreviation | Definition or Meaning |
|---|---|
| CP | child pornography |
| O | Offender |
| P | Primary |
| yrs | years |
| HH | household |
| f2f | face to face |
| p2p | peer to peer |
| wk | week |

**Technical support for this dataset is provided by NDACAN.**

**Please send your inquiries to NDACANSUPPORT@cornell.edu**