

Cross-Tabulations and Units of Count
with
NIBRS Data Elements

Programs Support Section
Criminal Justice Information Services Division
Federal Bureau of Investigation
October 1997

Introduction

The National Incident-Based Reporting System (NIBRS) is an incident-based crime reporting program for local, state, and federal law enforcement agencies. For each incident, NIBRS captures information on offenses, victims, offenders, arrestees, and properties involved and provides linkages among them. This wealth of information requires a degree of sophistication to analyze and utilize such data. A description of this program and its reporting requirements are beyond the scope of this paper. Detailed particulars of NIBRS can be found in the References at the end of this paper. Reference [1] provides a general description or overview, References [2] to [6] details, and References [7] and [8] its utilization. This paper is an expansion of the paper by Dr. Yoshio Akiyama, “Feasibility of Cross-Tabulations Illustrated Through NIBRS Data Elements”, August 1992.

NIBRS allows various dimensions of analysis including incident-based, offense-based, and victim-based analyses, which are addressed in Reference [7]. The related topic “units of count”, associated with these analyses, needs to be more fully addressed here. While NIBRS is most frequently characterized by its comprehensive structure of 53 data elements, not every given set of NIBRS variables (or data elements) are immediately cross-tabulable. Therefore, the object of this report is to address the feasibility of cross-tabulations among NIBRS variables and the associated units of count. When a cross-tabulation is not feasible, solutions will be suggested. This paper will not cover all logical cases but instead highlights meaningful, frequently encountered cross-tabulations.

Several points need to be kept in mind while reading this paper:

1. While this paper addresses NIBRS exclusively, the principles for statistical analyses discussed herein have general applicability. In fact, any database with structures similar to NIBRS can be analyzed in a similar manner.
2. The principles described herein pertain to “first level” cross-tabulations from which more refined analyses can be made. For example, while the feasibility of cross-tabulating the “date stolen” and the “date recovered” variables is discussed, the distribution of the time lapse between theft and recovery, which can be viewed as a derivative from this first level cross-tabulation, is not discussed here.
3. It is a common practice to “recode” or “regroup” data values of variable(s) into “blocks/chunks” which are more suitable/meaningful for one’s analytical purposes. An example is given below.

Example: Instead of listing data by single age categories, offenders in an incident are sometimes grouped into the categories “Adults Only”, “Juveniles Only”, “Adults and Juveniles”, and “Unknown”.

Discussion in this paper will involve the “raw” data not “recoded/regrouped” data as illustrated in the preceding Example. However, the analytical methodologies and units of count used for both data “types” are identical.

4. Skipping certain NIBRS variables in the chain of linkages as presented in the diagram in Section I and cross-tabulating variables not adjacent to each other (due to skipplings) could be a cause for information loss and provide “different” result. This is explained below by the following NIBRS example.

Example: Consider the situation where the variable X (Incident Number -- Data Element #2) is cross-tabulated with the variable Z (Type Injury -- Data Element #33) by skipping the variable Y (Victim Sequence Number -- Data Element #23) using the following hypothetical incident.

Incident Number (variable X) 0000001 has three victims with injuries as shown below -

Victim Sequence # (Y)	Type Injury (Z)
Victim 1	Unconsciousness, Broken Bones
Victim 2	Broken Bones
Victim 3	Internal Injury

If Y is taken into account, we would have the following table:

Type Injury (Z)	# of Victims (Y)	# of Incidents (X)
Unconsciousness	1	1
Broken Bones	2	1
Internal Injury	1	1

However, if Y is skipped, then only incident counting via X is feasible as shown in the table below:

Type Injury (Z)	# of Incidents (X)
Unconsciousness	1
Broken Bones	1
Internal Injury	1

I. The Structure of NIBRS Data

The 53 data elements comprising NIBRS and the six Group “A” Incident Report Segments to which they belong are listed below. The first column provides the data element number, and the second column provides the corresponding data element name:

Administrative Segment

- | | |
|---|----------------------------|
| 1 | ORI Number |
| 2 | Incident Number |
| 3 | Incident Date/Hour |
| 4 | Cleared Exceptionally |
| 5 | Exceptional Clearance Date |

Offense Segment

- | | |
|----|--------------------------------|
| 1 | ORI Number |
| 2 | Incident Number |
| 6 | UCR Offense Code |
| 7 | Offense Attempted/Completed |
| 8 | Offender(s) Suspected of Using |
| 8A | Bias Motivation |
| 9 | Location Type |
| 10 | Number of Premises Entered |
| 11 | Method of Entry |
| 12 | Type Criminal Activity |
| 13 | Type Weapon/Force Involved |

Property Segment

- | | |
|----|------------------------------------|
| 1 | ORI Number |
| 2 | Incident Number |
| 14 | Type Property Loss/Etc. |
| 15 | Property Description |
| 16 | Value of Property |
| 17 | Date Recovered |
| 18 | Number of Stolen Motor Vehicles |
| 19 | Number of Recovered Motor Vehicles |
| 20 | Suspected Drug Type |
| 21 | Estimated Drug Quantity |
| 22 | Type Drug Measurement |

Victim Segment

1	ORI Number
2	Incident Number
23	Victim (Sequence) Number
24	Victim Connected to UCR Offense Code(s)
25	Type (of Victim)
26	Age (of Victim)
27	Sex (of Victim)
28	Race (of Victim)
29	Ethnicity (of Victim)
30	Resident Status (of Victim)
31	Aggravated Assault/Homicide Circumstances
32	Additional Justifiable Homicide Circumstances
33	Type Injury
34	Offender Number(s) to be Related
35	Relationship(s) of Victim to Offender(s)

Offender Segment

1	ORI Number
2	Incident Number
36	Offender (Sequence) Number
37	Age (of Offender)
38	Sex (of Offender)
39	Race (of Offender)

Arrestee Segment

1	ORI Number
2	Incident Number
40	Arrestee (Sequence) Number
41	Arrest (Transaction) Number
42	Arrest Date
43	Type of Arrest
44	Multiple Arrestee Segments Indicator
45	UCR Arrest Offense Code
46	Arrestee Was Armed With
47	Age (of Arrestee)
48	Sex (of Arrestee)
49	Race (of Arrestee)
50	Ethnicity (of Arrestee)
51	Resident Status (of Arrestee)
52	Disposition of Arrestee Under 18

In this paper, the terms “data elements”, “variables”, and “fields” are used

interchangeably. For discussion purposes, the 53 data elements are classified into two main groups -- identifying and observational variables. The observational variables are conventionally divided into three subgroups according to their data types and usages -- ratio, interval, and nominal variables.

Out of the 53 data elements, the following six data elements are used for referencing or linking purposes:

- 1 ORI Number
- 2 Incident Number
- 23 Victim (Sequence) Number
- 36 Offender (Sequence) Number
- 40 Arrestee (Sequence) Number
- 41 Arrest (Transaction) Number

For example, the ORI and Incident Numbers are used to link the six NIBRS Group “A” segments to form an incident-based system. In this paper, they will be referred to as “identifiers” or “identifying variables”.

The remaining 47 variables provide observational or attribute values and will be referred to as “observational” variables. They are commonly divided into three subgroups based on their data types and usages. The first subgroup consists of eight ratio or comparison variables:

- 10 Number of Premises Entered
- 16 Value of Property
- 18 Number of Stolen Motor Vehicles
- 19 Number of Recovered Motor Vehicles
- 21 Estimated Drug Quantity
- 26 Age (of Victim)
- 37 Age (of Offender)
- 47 Age (of Arrestee).

The second subgroup consists of four interval or time variables:

- 3 Incident Date/Hour
- 5 Exceptional Clearance Date
- 17 Date Recovered
- 42 Arrest Date

The third and last subgroup consists of the remaining 35 variables which are assigned nominal or categorical values.

Given two NIBRS variables X and Y where the value(s) of Y depends on the value of X,

i.e., Y is an attribute or a description of X, there are two possibilities in terms of the number of values that Y can take for each entry of X.

In the first possibility, the variable Y has a unique value for each entry of X (Y is single-valued). An example can be obtained from the Arrestee Segment by referring to Arrestee Sequence Number (Data Element #40) as X and Age of Arrestee (Data Element #47) as Y. Each arrestee has a unique age value reported. In this paper, the notations $X \rightarrow Y$ or $Y \leftarrow X$ will be used to denote this “single-valued” relationship to illustrate a general statement/principle arising from these type of variables. For the aforementioned example, the notations

Arrestee Seq # (Data Element #40) \rightarrow Age of Arrestee (Data Element #47)

or

Age of Arrestee (Data Element #47) \leftarrow Arrestee Seq # (Data Element #40)

(more simply, $40 \rightarrow 47$ or $47 \leftarrow 40$) may also be used.

In the second possibility, the variable Y is allowed to have multiple responses for an entry of X (Y is multiple-valued for one entry of X). An example can be obtained from the Offense Segment by referring to the UCR Offense Code (Data Element #6) as X and Type Weapon/Force Involved (Data Element #13) as Y. In this example, Y can have up to three values for an entry of X. More specifically, suppose three robbers held up a bank. One was armed with a revolver, the second had a sawed-off shotgun, and the third had an automatic machine gun. Then, for this case, UCR Offense Code (Data Element #6) is coded as “120” (Robbery) and Type Weapon/Force Involved (Data Element #13) takes the values “12” (handgun), “14” (shotgun), and “15A” (automatic other firearm) in the Offense Segment. This “multiple-valued” relationship between X and Y is denoted by $X \multimap Y$ or $Y \ni X$. In the example provided here, the notations $6 \multimap 13$ or $13 \ni 6$ may be used.

Since available data is the basis of this paper, no distinction needs to be made of the response requirements (i.e., mandatory, conditional, or optional) of a variable. In the example $6 \multimap 13$ mentioned above, Data Element #13 is a conditional variable since it is required only for specific UCR offense codes. For example, it is required for the offense of “Murder and Nonnegligent Manslaughter” but not for the offense of “Embezzlement”.

The following example illustrates the use of the pictorial structure of NIBRS that follows. Consider the “path” shown below which is a part of the pictorial diagram:

PROPERTY SEGMENT

2 (Incident Number) ————— € 14 (Type Property Loss/Etc.) —€ 15 (Property Description) —→ 16 (Value of Property)

This conveys that each “Incident Number” could have more than one Property Segment record. The allowed entries for “Type Property Loss/Etc.” are:

- 1 = None
- 2 = Burned
- 3 = Counterfeited/Forged
- 4 = Destroyed/Damaged/Vandalized
- 5 = Recovered
- 6 = Seized
- 7 = Stolen/Etc.
- 8 = Unknown

Thus the relationship **2** —€ 14 results. Data element numbers which are in bold and underscored are identifiers.

Furthermore, for each “Type Property Loss/Etc.”, there may be up to ten (10) property descriptions in “Property Description” field, each of which, in turn, has only one associated value, “Value of Property”. The former gives rise to the relationship 14 —€ 15 and the latter 15 —→ 16. The allowed entries for “Property Description” are as follows:

- 01 = Aircraft
- 02 = Alcohol
- 03 = Automobiles
- 04 = Bicycles
- 05 = Buses
- ⋮
- 77 = Other
- 88 = Pending Inventory
- 99 = Special category to be used by the National UCR Program to compile statistics on certain designated types of property, e.g., “CB” radios.

The “Value of Property” is reported in whole dollars with the value “1” being used to designate that the property is of unknown value.

Using the notations just established, the structure of the 53 NIBRS variables is shown in the figure on the next page, which describes pictorially the relationships among the 53 variables and the six NIBRS Segments. This structure forms the basis used to discuss the statistical capability in NIBRS. The diagram in the figure provides the complete relationships among the 53 NIBRS data elements and the associated six Group “A” Offenses segments as described in References [1] and [2].

II. Units of Count

The “unit of count” is important in analyzing NIBRS data. Consider, for example, the analysis of firearms used in offenses as illustrated in the table below.

Offense Category	Firearm*	Handgun	Rifle	Shotgun	Other Firearm
Homicide offenses:					
Murder & nonneg manslaughter	1	19	4	11	1
Negligent homicide	0	2	0	0	0
Justifiable homicide	0	2	0	1	0
Kidnapping/abduction	0	18	6	5	1
Assault offenses:					
Aggravated assault	88	758	166	230	30
Simple assault	0	0	0	0	0
Robbery	10	79	3	5	1
Sex offenses, forcible:					
Forcible rape	2	6	2	0	0
Forcible sodomy	0	1	0	0	0
Sexual assault with an object	0	0	0	0	0
Forcible fondling	0	0	0	0	0

* “Firearm” is used when its type is not stated.

This table indicates the types of firearm weapons associated with a given offense. It should be noted that the table entry figures do not represent the number of weapons. The figure “4” at the intersection of the row “Murder and nonnegligent manslaughter” and the column “Rifle” does not mean four (4) rifles were used, but rather four (4) murder and nonnegligent manslaughter offenses involved the presence of rifle(s).

It is noted from the Structure of NIBRS Data Elements that the relationship

UCR Offense Code (Data Element #6) —€ Type Weapon/Force Involved (Data Element #13)

holds in the above table. The unit of count (here, the number of offenses) arises from Data Element #6. Generally, it will be observed that the unit of count is inherent in the variable from which “—>” or “—€ ” originates. This is discussed further in Section IV. Additional Principles in Cross-Tabulations.

III. Cross-Tabulations and Units of Count

This section examines the structure of the NIBRS data described in Section I to arrive at the basic NIBRS statistical capability. The general guide is to trace the NIBRS data structural paths depicted there. Up to three variables are discussed here. The simplest case of one variable (identifier or observational) is addressed first. Then the case of two variables (two identifiers, one identifier and one observational variable, and two observational variables) is analyzed. Finally, three variables are presented. Guidelines for extensions to more variables are also provided. Examples are provided as motivations to arrive at the more general statements or cases.

III.1. One Variable Analyses

Although a one-variable situation does not give rise to cross-tabulations, it is described here to introduce notations and provide familiarity to the subsequent cross-tabulation discussion. This subsection discusses analyses involving one variable where Case 1 treats one identifier and Case 2 deals with one observational variable.

Case 1. Consider the simple case of one NIBRS identifier. As an example, consider the identifier “Victim Sequence Number (Data Element #23)”. Then “Victim” is the inherent unit of count. The following is a table consisting of hypothetical data from three incidents:

Incident #	Victim Seq #	# of Victim Seq Number Entries
001	1, 2, 3	3
002	1, 2	2
003	1, 2, 3, 4	4
Total	N/A	9 victims

Some notations must be established for use in the resulting General Statement. For the preceding example, let

I = the identifier “Victim Sequence Number (Data Element #23)”,

[I] = the inherent unit of count - “Victim”,

and

||I|| = the number of entries referenced by I.

Then the third column total “9 victims” is the sum of the ||I||s.

Case 1 General Statement:

For an identifier I , let $[I]$ be the unit of count inherent in I and $\|I\|$ be the number of entries referenced by I . Then, a statistical output is the sum of the $\|I\|$ s.

Case 2. Next, consider the case of one observational variable. As an example, consider the observational variable “Sex of Offender (Data Element #38)”. Then “Offender” is the inherent unit of count. The following is a hypothetical table of the frequency distribution of offenders based on sex:

Sex of Offender	# of Offenders (frequency)
Male	170
Female	52
Total	222

Case 2 General Statement:

Each observational variable gives rise to a frequency distribution of values taken by that variable with the unit of count also inherent in that same variable.

Note: The above general statement is not meaningful for “Estimated Drug Quantity (Data Element #21)”, since it in itself offers only incomplete information unless combined with the observational variable “Type Drug Measurement (Data Element #22)”. This is strictly due to the NIBRS design that Data Elements #21 and #22 in combination can offer a complete information. Sometimes it takes more than one NIBRS variable to produce a meaningful observational variable.

III.2. Two-Variable Analyses

Cross-tabulation of two variables are addressed below. The possible combination of two variables and the cases which address them are enumerated below:

- For two identifiers I_1 and I_2 there are two possible relationships:

$$I_1 \rightarrow I_2 \text{ (See Case 3)}$$

$$I_1 \in I \text{ (See Case 4).}$$

- For an identifier I and an observational variable O there are two possible relationships:

$$I \longrightarrow O \text{ (See Case 5)}$$

$$I \text{ ---} \in O \text{ (See Case 6).}$$

- For two observational variables O_1 and O_2 there are two possible relationships:

$$O_1 \longrightarrow O_2 \text{ (See Case 7)}$$

$$O_1 \text{ ---} \in O_2 \text{ (See Case 8).}$$

Note: For an identifier I and an observational variable O it is not possible in NIBRS to have the following relationships:

$$O \longrightarrow I \text{ and } O \text{ ---} \in I.$$

Case 3. Consider the case of two NIBRS identifiers I_1 and I_2 satisfying the relationship:

$$I_1 \longrightarrow I_2.$$

Note that the relationship $I_1 \longrightarrow I_2$ is nothing more than a renaming of the identifiers and generates the same output as in Case 1. Either $[I_1]$ or $[I_2]$ can serve as the appropriate unit of count. An examination of the NIBRS structure in Section I shows that this relationship only exists between the two identifiers “Arrestee Sequence Number (Data Element #40)” and “Arrest Transaction Number (Data Element #41)”.

Case 3 General Statement:

Consider two NIBRS identifiers I_1 and I_2 satisfying the relationship $I_1 \longrightarrow I_2$. Since the relationship $I_1 \longrightarrow I_2$ is nothing more than a renaming of the two identifiers, it generates the same output as in Case 1. Either $[I_1]$ or $[I_2]$ can serve as the appropriate unit of count.

Case 4. Consider the case of two NIBRS identifiers I_1 and I_2 satisfying the relationship:

$$I_1 \text{ ---} \in I_2.$$

As an example, let I_1 be the identifier “ORI Number (Data Element #1)” and let I_2 be the identifier “Incident Number (Data Element #2)”. The relationship between them is indeed $I_1 \text{ ---} \in I_2$. Using $[I_1]$ as the unit of count, the following hypothetical frequency distribution table can be obtained for 15 imaginary agencies:

# of Incidents ($\ I_2\ $)	# of ORIs (frequency)
100	5
159	2
200	8
Total	15

Case 4 General Statement:

Let I_1 and I_2 be two identifiers with the relationship $I_1 \text{ ---} \in I_2$. Using $[I_1]$ as the unit of count, the following frequency distribution table can be obtained:

$\ I_2\ $	Frequency
1	F_1
2	F_2
3	F_3
\vdots	\vdots
Total	F

Up to this point only two identifiers have been discussed. At this juncture, the discussion extends to an identifier and an observational variable. There are two possible relationships between these two variables:

Identifier \rightarrow Observational (See Case 5)

Identifier $\text{---} \in$ Observational (See Case 6).

Case 5.

Let I be the identifier “Offender Sequence Number (Data Element #36)” and O be the observational variable “Sex of Offender (Data Element #38)”. O has the following admissible values:

$$O_1 = \text{“Male”}, O_2 = \text{“Females”}, \text{ and } O_3 = \text{“Unknown”}.$$

Then, the unit of count is [I] = “Offender” and the relationship between these two variables is $I \rightarrow O$ or, equivalently, $36 \rightarrow 38$. The following hypothetical statistical output is presented:

Offender’s Sex	Offender Frequency
Male	90
Female	20
Unknown	3
Total	113

Case 5 General Statement:

Let I be an identifier and O be an observational variable with admissible values O_1, O_2, O_3, \dots satisfying the relationship $I \rightarrow O$. Using [I] as the unit of count, the following statistical output can be obtained:

Observations	Frequency
O_1	F_1
O_2	F_2
O_3	F_3
\vdots	\vdots
Total	F

Case 6.

Let I be the identifier “Victim Sequence Number (Data Element #23)” and O be the observational variable “Victim Connected to UCR Offense Codes (Data Element #24)”. Each victim can be associated with up to 10 offense codes and hence we have the relationship $I \in O$ or, equivalently, $23 \in 24$. Using [I] = “Victim” as the unit of count, we may have the following hypothetical output:

Offense Code	# of Victims
Arson	3
Aggravated Assault	7
Simple Assault	8
Intimidation	12
⋮	⋮

Note: Summing the second column is not meaningful because it does not give the total number of victims reported due to duplicative counting by noting that multiple offenses can be inflicted on a given victim. If this table were restricted to single offense incidents, then the second column total makes sense.

Case 6 General Statement:

Let I be an identifier and O be an observational variable with admissible values O_1, O_2, O_3, \dots satisfying the relationship $I \text{ ---} \in O$. Then the following statistical output can be obtained with the unit of count being $[I]$:

Observations O_j	Frequency
O_1	F_1
O_2	F_2
O_3	F_3
⋮	⋮

Due to the duplicative countings in F_j , the total of the F_j 's is not meaningful.

Note: It is, however, possible to restrict the database to “single-response” situations (e.g., “single-offense” incidents) or institute a hierarchy scheme to choose a single (unique) answer out of multiple responses. In such tables the second column totals are meaningful.

To complete the discussion of two variables, two observational variables will now be considered. There are two possible relationships between two observational variables, namely,

Observational 1 \rightarrow Observational 2 (See Case 7)

Observational 1 $\text{---} \in$ Observational 2 (See Case 8).

Case 7. Consider the example where the two observational variables are $O =$ “Property

Description (Data Element #15)” and $O' = \text{“Date Recovered (Data Element #17)”}$. Then the relationship $O \rightarrow O'$ or, equivalently, $15 \rightarrow 17$ holds and a unique response is provided only for those properties that were recovered. A two-dimensional cross-tabulation can be generated for these two variables where the units of count are “aircrafts”, “bicycles”, etc. as specified in Data Element #15. The following is a hypothetical table.

Date Recovered	Aircraft	Bicycle	Automobile	Radio/TV/ VCR	Total
3/10/97	0	10	4	20	34
4/6/97	2	5	8	0	15
6/3/97	0	29	0	15	44
6/22/97	0	0	10	3	13
Total	2	44	22	38	106

Case 7 General Statement:

Let O and O' be two observational variables with the relationship $O \rightarrow O'$. Denote the values of the variables O and O' by O_j and O'_i , respectively. Then a two-dimensional cross-tabulation can be generated using the unit of count, $[O]$, as illustrated:

	...	O_j	...	Total
\vdots		\vdots		\vdots
O'_i	...	$F(i, j)$...	$F(i, \cdot)$
\vdots		\vdots		\vdots
Total	...	$F(\cdot, j)$...	F

$F(i, j)$ is the number of O_j 's given O'_i . $F(i, \cdot)$ is the i -th row total and $F(\cdot, j)$ is the j -th column total.

Case 8. Consider the example where the observational variables $O = \text{“UCR Offense Code (Data Element #6)”}$ and $O' = \text{“Offenders Suspected of Using (Data Element #8)”}$. They satisfy the relationship $O \in O'$ or, equivalently, $6 \in 8$. Using the unit of count $[O] = \text{“Offense,”}$ the following hypothetical table can be constructed for

murder and aggravated assault:

Offenders Suspected of Using	Murder	Aggravated Assault	Total
Alcohol	2	9	11
Drug	10	17	27
Computer	0	2	2
None	8	15	23

The hypothetical figure “11” in the total column represents the number of offenses that involved the usage of alcohol. Vertical totals are not meaningful due to duplicative countings of offenses. For example, an offense could involve the usage of both alcohol and drug. Contrast this with example given in Case 7.

One should note, however, that if “Offenders Suspected of Using” is subjected to a finer breakdown (to avoid duplicative counting) as shown in the next table, then vertical sums would be meaningful.

Offenders Suspected of Using	Murder	Aggravated Assault	Total
Alcohol Only	1	7	8
Drug Only	9	15	24
Computer Only	0	0	0
Alcohol & Drug	1	0	1
Alcohol & Computer	0	0	0
Drug & Computer	0	0	0
Alc, Drug & Comp	0	2	2
None	8	15	23
Total	19	39	58

Case 8 General Statement:

Let O and O' be two observational variables with the relationship $O \text{ —} \in O'$. Denote the values of the variables O and O' by O_j and O'_i , respectively. Using $[O]$ as the unit of count the following table can be constructed:

	... O _j ...	Total
⋮	⋮	⋮
O' _i	... F(i, j) ...	F(i, ·)
⋮	⋮	⋮

No vertical totals are meaningful due to duplicative countings in contrast to Case 7.

Note: By refining O' so that no duplication is present, one can make a column-summable table as explained in the above example.

III.3. Three-Variable Analyses

To complete this section, the examination of the NIBRS data element structure for statistical capability is now extended to three variables. The possible combinations are listed below:

- Three identifiers
- Two identifiers and one observational variable
- One identifier and two observational variables
- Three observational variables

The last three combinations will be addressed below (Cases 9-17) as they represent more interesting and meaningful cases.

Two identifiers and one observational variable will be considered first. From Case 3 we know that the only interesting relationship between two NIBRS identifiers is Identifier 1 $\text{---}\in$ Identifier 2. Consider the following two relationships involving an observational variable:

Identifier 1 $\text{---}\in$ Identifier 2 $\text{---}\rightarrow$ Observational (Case 9)

Identifier 1 $\text{---}\in$ Identifier 2 $\text{---}\in$ Observational (Case 10).

Case 9. Let I be the identifier "Victim Sequence Number (Data Element #23)", I' be the identifier "Incident Number (Data Element #2)", and O be the observational variable "Cleared Exceptionally (Data Element #4)". Then the relationship I $\text{---}\in$ I' $\text{---}\rightarrow$ O is satisfied. However, if ||I||, the number of victims involved in an incident, is considered as a variable then we have the relationship ||I|| $\text{---}\leftarrow$ I' $\text{---}\rightarrow$ O.

The following table can then be generated for the latter where the unit of count is “Incidents”:

Data Element #4	# of Victims Involved, I				Total
	1	2	3	...	
Death of Offender	7	
Prosecution Declined	.				
Extradition Denied	.				
Victim Refused to Cooperate	.				
Juvenile/No Custody	.				
Not Applicable	.				
Total					

The hypothetical number “7” in the above table, for instance, means that there were 7 incidents that involved only one victim and were exceptionally cleared by the death of offender(s).

Case 9 General Statement:

Let I and I' be identifiers and O be an observational variable satisfying the relationship $I \ni - I' \rightarrow O$. A cross-tabulation can be generated by ||I|| and O with the unit of count being [I']. All vertical and horizontal totals are meaningful.

Case 10. Consider the example where I is the identifier “Offender Sequence Number (Data Element #36)”, I' is the identifier “Incident Number (Data Element #2)”, and O is the observational variable “UCR Offense Code (Data Element #6)”. Then the relationship $I \ni - I' \rightarrow O$ is satisfied. Replacing I with ||I||, the number of offenders in a given incident, we have the relationship $||I|| \leftarrow I' \rightarrow O$ from which a statistical table can be generated where the unit of count used is [I'] = “Incidents”.

Offense (Data Element #6)	# of Offenders per Incident, $\ I\ $		Total
	...	5	...
⋮		⋮	⋮
Murder	...	20	...
⋮		⋮	⋮

The entry “20” means that there were 20 incidents where there were five offenders in each incident and murder was one of the offenses committed. Vertical sums are not meaningful due to the possibility of multiple offenses in an incident. Each horizontal sum gives the total number of incidents involving at least that particular offense.

Case 10 General Statement:

Let I and I' be two identifiers and O be an observational variable with admissible values O_i satisfying the relationship $I \ni I' \in O$. The following table can be generated from the derived relationship $\|I\| \leftarrow I' \in O$ where the unit of count used is $[I']$.

	$\ I\ $		Total
	...	j	...
⋮		⋮	⋮
O_i	...	$F(i, j)$...
			$F(i, \cdot)$
⋮		⋮	⋮

Note: The portion $I' \in O$ shows that multiple responses are allowed for O . Therefore, vertical sums in the above table would not be meaningful. The horizontal sums (in the right column), however, are meaningful.

Extending the discussion of three variables, Cases 11, 12, and 13 will address the statistical analysis of one identifier and two observational variables. The relationships considered are:

Observational 1 \leftarrow Identifier \rightarrow Observational 2 (Case 11),

Observational 1 \ni — Identifier \rightarrow Observational 2 (Case 12),

and

Observational 1 \ni — Identifier \notin Observational 2 (Case 13).

Case 11. Let O be the observational variable “Age of Arrestee (Data Element #47)”, O’ be the observational variable “Sex of Arrestee (Data Element #48)”, and I be the identifier “Arrestee Sequence Number (Data Element #40)”. Then the relationship $47 \leftarrow 40 \rightarrow 48$ or $O \leftarrow I \rightarrow O'$ holds. The arrestees are tabulated by age and sex with the unit of count being [I] = “Arrestee”. The table presented below is similar to that in Case 7 except for the choice of the unit of count which was [O] in Case 7.

Sex of Arrestee	Age					Total
	...	23	24	25	...	
Male						
Female						
Total						

Case 11 General Statement:

Let I be an identifier and let O, O’ be two observational variables satisfying the relationship $O \leftarrow I \rightarrow O'$. Then, statistical output same as Case 7 can be generated with [I] being the unit of count.

Case 12. Let O be the observational variable “Arrestee was Armed With (Data Element #46)”, O’ be the observational variable “Type of Arrest (Data Element #43)”, and I be the identifier “Arrest Sequence Number (Data Element #40)”. Then the relationship $46 \ni$ — $40 \rightarrow 43$ or $O \ni$ — $I \rightarrow O'$ holds. This is similar to the example given in Case 11 except that columns are not summable due to duplicative listings.

Case 12 General Statement:

Let I be an identifier and let O, O’ be two observational variables satisfying the relationship $O \ni$ — $I \rightarrow O'$. Using the unit of count, [I], $O \ni$ — $I \rightarrow O'$ gives

rise to the following table:

	...	O'_j	...	Total
\vdots		\vdots		\vdots
O_i	...	$F(i, j)$...	$F(i, \cdot)$
\vdots		\vdots		\vdots

Due to the duplicative countings, column sum $F(\cdot, j)$ is not meaningful.

Case 13. This case takes care of the relationship:

Observational 1 \ni — Identifier \rightarrow Observational 2.

Case 13 General Statement:

Let I be an identifier and let O and O' be observational variables satisfying the relationship $O \ni$ — $I \rightarrow$ O' . Often this does not provide a meaningful cross-tabulation between O and O' . There are possible ways to improve this situation. Three such ways are given below:

- a. Limit the identifier variable I to I_1 where the response for O is unique. Then, the relationship $O \leftarrow I_1 \rightarrow O'$ will provide a statistical cross-tabulation as in Case 12.
- b. Limit the identifier variable I to I_2 where the response for O' is unique. Then, the relationship $O \ni$ — $I_2 \rightarrow O'$ will provide a statistical cross-tabulation as in Case 12.
- c. Limit the identifier variable I to I_3 so that the responses for O and O' are unique. Then the relationship $O \leftarrow I_3 \rightarrow O'$ holds and thus provides a statistical output as in Case 11.

Three observational variables are now considered. The possible non-equivalent relationships among them are as follows:

Observational 1 \leftarrow Observational 2 \rightarrow Observational 3 (See Case 14),

Observational 1 \ni — Observational 2 \rightarrow Observational 3 (See Case 15),

and

Observational 1 \ni — Observational 2 $\text{—}\in$ Observational 3 (See Case 16).

Case 14. Let O be the observational variable “Offense Attempted/Completed (Data Element #7)”, O' be the observational variable “UCR Offense Code (Data Element #6)”, and O'' be the observational variable “Location Type (Data Element #9)”. The relationship $O \leftarrow O' \rightarrow O''$ or $7 \leftarrow 6 \rightarrow 9$ holds and a three-way table in terms of these variables can be constructed with the unit of count being $[O'] =$ “Offense.” A table will be provided for each offense as illustrated below, i.e., an offense becomes the parameter of each table:

	Attempted	Completed
Air/Bus/Train Terminal		
Bank/Savings and Loan		
Bar/Night Club		
⋮		
Other/Unknown		

Case 14 General Statement:

Let O , O' , and O'' be three observational variables satisfying the relationship $O \leftarrow O' \rightarrow O''$. Then, a three-dimensional cross-tabulation in these three variables may be generated with the unit of count $[O']$ as illustrated in the example above.

Case 15. This case takes care of the following relationship for three observational variables:

Observational 1 \ni — Observational 2 \rightarrow Observational 3.

Case 15 General Statement:

Let O , O' , and O'' be three observational variables satisfying the relationship $O \leftarrow O' \text{—}\in O''$. Then, a three-dimensional cross-tabulation in O , O' , and O'' is possible with the unit of count $[O']$. Let $F(i, j, k)$ be the frequency count for O_i, O'_j, O''_k . Due to duplicative countings of O'' in the portion $O' \text{—}\in O''$, marginal totals $F(\cdot, \cdot, k)$, $F(\cdot, j, k)$, and $F(i, \cdot, k)$ are meaningful but $F(\cdot, j, \cdot)$,

$F(i, \cdot, \cdot), F(i, j, \cdot),$ and $F = F(\cdot, \cdot, \cdot)$ are not.

Case 16. The last relationship among three observational variables to be discussed is:

Observational 1 \exists — Observational 2 $\text{—}\in$ Observational 3

Case 16 General Statement:

Let $O, O',$ and O'' be observational variables satisfying the relationship $O \exists\text{—} O' \text{—}\in O''$. Imposing the restrictions $O \leftarrow O'_1 \text{—}\in O''$, $O \exists\text{—} O'_2 \rightarrow O''$, or $O \leftarrow O'_3 \rightarrow O''$, as was done in Case 13 would enable cross-tabulations using the unit of count [O'].

In Cases 11, 12, and 13, two observational variables emanating from one identifier were discussed. The case below (Case 17) discusses the following different relationship involving the same types of three variables:

Identifier $\text{—}\in$ Observational 1 \rightarrow Observational 2,

i.e.,

$I \text{—}\in O \rightarrow O'$ in notation.

Case 17. Let I be the identifier “Incident Number (Data Element #2)”, O be the observational variable “UCR Offense Code (Data Element #6)”, and O' be the observational variable “Offense Attempted/Completed (Data Element #7)”. Then the following table can be generated with the unit of count being [I] = “Incident”.

Offense Category	Completed	Attempted
Crimes Against Persons		
Assault Offenses	226,417	0
Homicide Offenses	654	0
Kidnaping/Abduction	1,626	246
Sex Offenses, Forcible	11,690	717
Sex Offenses, Nonforcible	1,077	19
Crimes Against Society		
Drug/Narcotic Offenses	67,098	421
Gambling Offenses	192	13
Pornography/Obscene Material	241	2
Prostitution Offenses	1,849	28
Weapon Law Violations	10,564	275
Crimes Against Property		
Arson	3,695	338
Bribery	10	11
Burglary/Breaking and Entering	108,067	9,427
Counterfeiting/Forgery	11,965	367
Destruction/Damage/Vandalism	175,035	701
Embezzlement	3,126	25
Extortion/Blackmail	96	49
Fraud Offenses	12,759	826
Larceny/Theft Offenses	406,262	7,589
Motor Vehicle Theft	41,444	2,581
Robbery	10,415	1,294
Stolen Property Offenses	5,482	32

Case 17 General Statement:

Let I be an identifier and let O and O' be two observational variables satisfying the relationship $I \in O \rightarrow O'$. Denote the admissible values of the variables O and O' by O_i and O'_j , respectively. Then a two-dimensional cross-tabulation can be generated using the unit of count, $[I]$, as illustrated:

	...	O'_j	...	Total
\vdots		\vdots		\vdots
O_i	...	$F(i, j)$...	$F(i, \cdot)$
\vdots		\vdots		\vdots

No vertical totals are meaningful due to duplicative countings.

IV. Additional Principles in Cross-Tabulations

Additional principles that are applicable in any data structure such as NIBRS are mentioned below. They may be used to extend the methods mentioned in the previous section where the discussion has been limited to no more than three variables.

1. The unit of count can be determined by imposing a hierarchical consideration. The following examples presented in the table below demonstrate this point.

Structure	Unit of Count
$X \rightarrow Y$	[X]
$N \text{---} \in Y$	[N]
$X \rightarrow Y \rightarrow Z$	[X]
$X \leftarrow Y \text{---} \in Z$	[Y]
$X \ni \text{---} N \rightarrow Y$	[N]

In $X \rightarrow Y$, for instance, X precedes Y. Therefore, the unit of count is [X]. In $X \leftarrow Y \text{---} \in Z$, Y precedes X and Z, hence the unit of count is [Y].

2. Up to this point, this discussion has been limited to chains with up to three variables. To extend this concept, the following table is provided to illustrate the reduction scheme of 3-chains into 2-chains. By iterations, this table can be used to reduce any chain into a 2-chain.

3-Chains	2-Chains
$X \rightarrow Y \rightarrow Z$	$X \rightarrow Z$
$X \rightarrow Y \text{---} \in Z$	$X \text{---} \in Z^*$
$X \text{---} \in Y \rightarrow Z$	$X \text{---} \in Z^*$
$X \text{---} \in Y \text{---} \in Z$	$X \text{---} \in Z^*$

*A stronger result may be obtained depending on the restriction placed.

For example, consider the Victim Segment with the variables “ORI Number (Data Element #1)”, “Incident Number (Data Element #2)”, “Victim Sequence Number (Data Element

#23”, “Offender Numbers to be Related (Data Element #34)”, and “Relationships of Victim to Offenders (Data Element #35)”. The relationships among these variables are as follows:

$$1 \text{ ---} 2 \text{ ---} 23 \text{ ---} 34 \text{ ---} 35.$$

This can first be reduced to

$$1 \text{ ---} 23 \text{ ---} 34 \text{ ---} 35,$$

then to

$$1 \text{ ---} 34 \text{ ---} 35,$$

and finally to

$$1 \text{ ---} 35.$$

As mentioned in the Introduction, skipping/reduction of intermediary variable(s) requires special attention to avoid invalid cross-tabulations.

3. If two variables X and Y are to be related, a common ramification/connection point must be found.

Example A: If the variable “Arrestee Was Armed With (Data Element #46)” is to be linked to the variable “Sex of Victim (Data Element #27)”, then the variable “Incident Number (Data Element #2)” is the ramification point as illustrated by:

$$27 \text{ ---} 25 \text{ ---} 23 \text{ ---} 2 \text{ ---} 40 \text{ ---} 46$$

where

23 is the variable “Victim Sequence Number (Data Element #23)”,
 25 is the variable “Type of Victim (Data Element #25)”, and
 40 is the variable “Arrestee Sequence Number (Data Number #40)”.

This long relationship sequence can then be reduced to $27 \text{ ---} 2 \text{ ---} 46$. Therefore, the analyses of 27 and 46 falls into Case 13.

Example B: If the variable “Location Type (Data Element #9)” is to be related to the variable “Sex of the Arrestee (Data Element #48)”, the relationship

$$9 \leftarrow 6 \ni 2 \in 40 \rightarrow 48$$

where

6 is the variable “UCR Offense Code (Data Element #6)” and
40 is the variable “Arrestee Sequence Number (Data Number #40)”

can be used. It is then reduced to $9 \ni 2 \in 48$. Therefore, the analysis must be conducted for either “single offense” incidents or “single arrestee” incidents, as mentioned in Case 13.

4. Two NIBRS variables belonging to two different NIBRS segments are linked only through the “Incident Number (Data Element 2).” This rule has two exceptions: Victims are related to offenses via “Victim Connected to UCR Offense Code(s) (Data Element #24)” and victims are related to offenders via “Offender Number(s) to be Related (Data Element #34)”. This is shown in the diagram in Section I.

References

1. Federal Bureau of Investigation, *Uniform Crime Reporting Handbook - NIBRS Edition*, 1992, Washington, D.C.
2. Federal Bureau of Investigation, *National Incident-Based Reporting System, Volume 1 Data Collection Guidelines*, 1996, Washington, D.C.
3. Federal Bureau of Investigation, *National Incident-Based Reporting System, Volume 2 Data Submission Specifications*, 1992, Washington, D.C.
4. Federal Bureau of Investigation, *National Incident-Based Reporting System, Volume 3 Approaches to Implementing an Incident-Based Reporting System*, 1992, Washington, D.C.
5. Federal Bureau of Investigation, *National Incident-Based Reporting System, Volume 4 Error Message Manual*, 1992, Washington, D.C.
6. Federal Bureau of Investigation, *National Incident-Based Reporting System, Supplemental Guidelines for Federal Participation*, 1990, Washington, D.C.
7. Federal Bureau of Investigation, *National Incident-Based Reporting System: Select Statistical Information Capabilities*, 1991, Washington, D.C.
8. Federal Bureau of Investigation, “*Utilization of UCR/NIBRS Data - A Discussion Paper*,” 1997, Washington, D.C.