## Attachment N. CASPER Toolkit, Sections 5.2 Weighted Analyses and 5.3

# **Calculation of 95% Confidence Intervals**

(Complete toolkit available at

http://emergency.cdc.gov/disasters/surveillance/pdf/CASPER\_Toolkit\_Version\_2\_0\_508\_Compliant.pdf)

### 5.2 Weighted analyses

Households selected in cluster sampling have an unequal probability of selection. To avoid biased estimates, *all data analyses should include a mathematical weight for probability of selection*. Once all data are merged into a single electronic dataset, a weight variable must be added to each surveyed household by use of the formula below:

Weight = Total number of housing units in sampling frame (number of housing units interviewed within cluster)\*(number of clusters selected)

The sampling frame, referred to in the numerator, is defined as the entire assessment area in which CASPER is being conducted. The numerator is the total number of housing units in the sampling frame, and that number will be the same for every assessed household. To calculate the total number of houses in the sampling frame, follow the steps outlined in <u>Section 3.1.3</u> and sum the "housing units" column (e.g., 6292 houses in Caldwell County, Kentucky).

If sampling has been 100% successful and information was obtained from exactly seven households in exactly 30 clusters, the denominator will be 7 \* 30 = 210 for every housing unit. The sample, then, is self-weighting because all housing units in the sample had an equal probability of being selected. Likely, obtaining seven households in each of the 30 clusters will not be possible. When this occurs, the denominator will be different for each surveyed household, depending on the cluster from which the housing unit was selected. Households from the same cluster will have the same weight, but weights will differ between clusters. For example, if only five completed interviews occurred in a cluster, the denominator of the weight for each of the five surveyed households would be 5 \* 30 = 150.

The "number of clusters selected" will be 30, even if there are some clusters with zero interviews. The only exception is if the decision to oversample clusters was made a priori (see <u>Section 3.2</u>).

The table depicted in Figure 11 displays the sampling weights for a CASPER conducted in Kentucky following the major ice storms in 2009. In stage one of sampling, 30 clusters were selected representing 19,370 housing units. The goal was to conduct 210 interviews, but only 187 were completed. For the purpose of calculating the "weight" column (highlighted in yellow), an additional column was added, "# interviews," to represent the number of housing units interviewed within a cluster (highlighted in blue).

	Tables		Illustrati	ons	Charts							5 Sparklines		
	D2	•	f <sub>x</sub>	=(1937	0/(C2*30))									
	A	В	С	D	E	F	G	Н		J	K	L	М	N
1	Date	Cluster_ID	# Interviews	Weight	Q2PeopleLi	Q3PeopleSI	aTwo	b3to17	c18to64	d65	Q4safe	Q5secure	Q6Injured	N6aFall
2	2/7/2009	2	6	107.61	4	6		2	3	1	yes	yes	no	no
3	2/7/2009	2	6	107.61	3	2			2		yes	yes	no	no
4	2/7/2009	2	6	107.61	1	1				1	yes	yes	no	no
5	2/7/2009	3	6	107.61	2	2			1	1	yes	yes	no	no
6	2/7/2009	3	6	107.61	2	2			1	1	yes	yes	no	no
7	2/7/2009	3	6	107.61	2	1			1		yes	yes	no	no
8	2/7/2009	4	7	92.24	2	2	0	0	0	2	yes	yes	no	no
9	2/7/2009	4	7	92.24	2	2	0	0	2	0	yes	yes	no	no
10	2/7/2009	4	7	92.24	2	2	0	0	2	0	yes	yes	no	no
11	2/7/2009	4	7	92.24	2	2	0	0	0	2	yes	yes	no	no
12	2/7/2009	4	7	92.24	5	6	0	2	4	0	yes	yes	no	no
13	2/7/2009	4	7	92.24	6	8	0	4	4	0	yes	yes	yes	yes
14	2/7/2009	4	7	92.24	4	7	0	0	7	0	yes	yes	no	no
15	2/7/2009	5	6	107.61	5	14	1	8	5		yes	yes	no	no
16	2/7/2009	5	6	107.61	4	4		2	3		yes	yes	no	no
17	2/7/2009	5	6	107.61	8	8		5	2	1	yes	yes	no	no
18	2/7/2009	5	6	107.61	3	3			3		yes	yes	no	no
19	2/7/2009	5	6	107.61	2	2				2	yes	yes	no	no
20	2/7/2009	5	6	107.61	2	2			1	1	yes	yes	no	no
21	2/7/2009	10	5	129.13	2	1	0	0	0	1	yes	yes	no	no
22	2/7/2009	10	5	129.13	2	2	0	0	0	2	yes	yes	no	no
0.0	2/7/2000	10	-	100 10	4	4	0	0	4	0				

Figure 11. Sample dataset showing the number of interviews per cluster and the assigned weight for each house interviewed.

Once weights are assigned, frequencies can be calculated for each of the interview questions. To calculate frequencies in Epi Info<sup>™</sup> 7 "classic mode", read (import) the data file with the weight that was just created. Click on "Frequencies" along the left hand column. In the "frequency of" box, select each variable for which you would like results and, in the "weight" box, select the variable "WEIGHT" that was just created. Finally, click "OK" (Figure 12) and a report will be generated providing the estimates. Figure 12. Epi Info<sup>™</sup> 7 "classic mode" frequency analysis window showing selected variables and weight.

37     65%    AII (") Except       Total 57     100%     Q20_ELECTR Q30_HH_ILL Q31_HH_INJ Q34_MEDNEE	L Freq %	Frequency of Stratify by
Meight Q20_ELECTR   Weight Q30_HH_ILL   Q31_HH_INJ   Q34_MEDNEE	- 37 65%	
Weight Q20_ELECTR Q30_HH_ILL Q31_HH_INJ Q34_MEDNEE	otal 57 100%	
Jutput to Table	eight /eight "tput to Table	Q20_ELECTR Q30_HH_ILL Q31_HH_INJ Q34_MEDNEE

Figure 13 displays the Epi Info<sup>™</sup> output window with the selected variables, followed by a table

for each selection. These output tables should be saved for use in the report.

Epi Into 7 - Analysis							
File View Tools Help							
mmand Explorer	P Output:output4.ht	ml					
Commands	Previous N	ext 🕪 Last 🚇 History	0pen	👚 Bookmark 🛛 🛲	Print		
	EDEO OM E	LECTD O20 HE	пто	21 HH INL	024 MEDN	FE WEICHTVAD	- 11
- 🗐 Relate		LECTR Q30_HF	I_ILL Q	SI_HH_INJ	Q34_MEDN	EE WEIGHIVAK	. = n
···· 🗎 Write (Export)							
🗎 Merge							
Delete File/Table	O20 FLECTR	Froquency	Porcont	Cum Porcont			
Undelete Records	v	17011 7702650700	1 Creent	00 220/			
🖃 🗁 Variables	Y	1/011.//93650/99	88.52%	88.32%			
Define	N	2250.60952380958	3 11.68%	100.00%			
Undefine	Total	19262.3888888889	100.00%	100.00%			
Assign							
🗎 Recode	95% Conf Li	mits					
Display	Y 87.85% 88	.76%					
Select/If	N 11.23% 12	.14%					
	Q30 HH ILL	Frequency	Percent	Cum. Percent			
	No	12584.3507936512	64.97%	64.97%			
Cancel Sort	Yes	6785.64920634949	35.03%	100.00%		1	
E Catistics	Total	19370.0000000007	100.00%	100.00%			
	95% Conf L	imits					
🗎 Means	No 64 29% 6	55 64%					
🗎 Summarize	Yes 34 36%	35 71%					
i im ⊫1 Granh	1 _103 5 1.5070 1					_	

Figure 13. Example of Epi Info<sup>™</sup> 7 "classic mode" output window showing weighted frequencies

To obtain unweighted estimates, follow the above instructions, but do not assign a variable in the "weight" box. Applying the weights provides projected estimates that can be generalized to every housing unit in the assessment area or sampling frame. <u>Table 7</u> shows the unweighted and weighted frequencies for a specific question from the 2009 Kentucky Ice Storm CASPER.

Table 7. Unweighted and weighted frequencies of current source of electricity following the Ice Storms, Kentucky, 2009

	Unweig	hted	Wei		
Characteristic	Frequency	Percent	Frequency	Percent	95% CI
Source of Electricity					
Power company	137	74.1	14190	74.0	61.9-86.0
Gasoline generator	29	15.7	3200	16.7	7.6-25.7
None	19	10.3	1789	9.3	3.8-14.8

Remember that weighted analysis does not account for the changes that may occur in the number of households between the time of the census and the time of the assessment (e.g., the number of households per cluster may have changed between 2000, when the census was conducted, and 2009, when the CASPER was conducted). Therefore, despite attempts to present unbiased estimates, the frequencies reported might lack precision.

### 5.3 Calculation of 95% confidence intervals

The 95% confidence intervals (CIs) should be provided with the weighted estimates. These confidence intervals indicate the reliability of the weighted estimate. Follow these steps to calculate 95% confidence intervals in Epi Info<sup>™</sup> 7:

1. Open Epi Info 7 in classic mode (Figure 14).



### Figure 14. Classic mode of Epi Info 7

2. Read (import) the data

file.

 Select "Complex Sample Frequencies Command" under advanced statistics, and in the dialog box for Frequency, select the variable(s) in which you are interested (Figure 15).

Figure 15 Selected variables for calculation of complex sample frequencies (sample data)											
LIGUING 15 NOIGCTON VARIANIAS FOR CAICULATION OF COMPLEY SAMPLE TRADUCING (SAMPLE DATA)		/ I		·····		· · · · · · ·		1 <b>f</b>	· · · · · · · · · · · · · · · · · · ·	C - I + I	
FIVILE 1.1. SEELED VALADUES DU LAUDUADUUT DITUUEL SAUDUE DEDUEDUES SAUDUE DALA?	ובזבח י	isamnie	eallencies	mnie trec	complex sar	ION OT	Calculation	ies tor	varianieg	Selected	olire 15
Tigure 13, Sciected variables for calculation of complex sample frequencies (sample data)		Jamph	cyuchicics	Inpic n cu	COMPICA Sal		calculation		variabic.	JUICTEU	SUIC IJ

Command Explorer 4							9 Y				
E- 2 Commands	Comple	ex San	ple Frequenc	les							
🖻 🦢 Data											
🚉 Read											
📄 Relate			1 m		F		- Ph. L.				
📄 Write (Export)	111	Frea	92		Frequency of	3	tratny by				
🔄 Merge			ru .								
Delete File/Table		20	35%			•	•				
Delete Records	+	20	5576								
Undelete Records		27	0.50								
Variables	_	31	65%								
Define					All (*) Except						
DefineGroup	Total	57	100%								
		•••			-						
Assign					Q20 ELECTR						
Dieslaw					020 1111 111						
Select //	Weigh	t –			Q30_HH_ILL						
Select		-			Q31 HH INJ						
Cancel Select	Weigh	ht		-	OOM MEDNICE						
	weigi	R.		· · )	Q34_MEDNEE						
Sort	D.:		Provide State								
Cancel Sort	<u>P</u> nma	ry Sai	npling Unit								
Statistics											
List		luster		•							
Frequencies											
🗐 Tables	Output	to Tak	le								
📖 Means	Juliput		NO								
🗐 Summarize											
Graph											
- Advanced Statistics											
🗐 Linear Regression											
🔄 Logistic Regression		ut				-					
Keplan Weler Sarvival	Se	ttings		OK	Cancel	Clear	Save Only Help				
Complex Sample Frequencies											
(a) Complex Sample Tables											
Complex Sample Means											

- 4. Under the Weight drop-down menu, select the "weight" variable for calculating the weighted CI.
- 5. Under PSU, select the "Cluster Number" variable and Click OK (Figure 16).



Figure 16. Example of 95% CI output in Epi Info<sup>™</sup> 7 "classic mode"

6. Right-click on the table and select "Export to Microsoft Excel".