

**BALANCING QUALITY AND CONFIDENTIALITY
IN PUBLIC HEALTH DATA**

Lawrence H. Cox, Ph.D.
National Institute of Statistical Sciences
cox@niss.org

ITSEW
Quebec City, Quebec
June 19-22, 2011

Public Health Data

- Vital statistics (birth, death, ...)
- Reported cases of (notifiable) communicable diseases
- Registry data
- Presented as counts by geography
 - State
 - County
 - Zip codeand by demographic variables
 - Gender
 - Age
 - Race/ethnicity
- Many two-way and often many-way tables
- Public use (microdata) files
- On-line (tabulation) query systems

Balancing Quality and Confidentiality

- Original data set is *masked* by confidentiality protection (*disclosure limitation = SDL*) methods, such as
 - suppression
 - rounding or perturbation
 - (controlled) adjustment
 - data swapping/switching
 - imputation/microaggregation/synthesization
- Disclosure limitation acts on original data, typically by
 - abbreviation
 - modification
 - replacement
- Thereby, SDL may affect aspects of data quality, such as
 - usability of data sets
 - accuracy, precision of individual data (local quality)
 - original data and error distributions (global quality)
 - inferences, decisions based on released masked data
- SDL is a component of TSE
- Currently, SDL is performed then readjusted to restore key quality features
- Better to combine quality-confidentiality considerations into a single methodology, as has been done for controlled tabular adjustment and rounding

Confidentiality Protection of Public Health Data

- Typically mandated by law, regulation, practice
- **Rule:** prohibit *small* counts (1 or 2)--but not in all cases
- **Problem:** variable but generally weak protection for the Rule--despite availability of proven methods

Today

- Illustrate the Problem via examples from published data
- Modify some descriptors, detail, etc. to anonymize
 - Subjects
 - Data releasers (Federal, State, Local)

REPORTED CASES, UNITED STATES, 2007

Area	Population (1000s)	AIDS†	Anthrax	Botulism			Brucellosis
				Foodborne	Infant	Other‡	
United States	299,398	37,503	1	32	85	27	131
New England	14,271	1,309	1 —	—	1 —	—	—
Connecticut	3,505	528	1 —	—	1 —	—	—
Maine	1,322	46 —	—	—	—	—	—
Massachusetts	6,437	612 —	—	—	—	—	—
New Hampshire	1,315	51 —	—	—	—	—	—
Rhode Island	1,068	66 —	—	—	—	—	—
Vermont	624	6 —	—	—	—	—	—
Mid. Atlantic	40,472	7,724 —	—	2	22	3	4
New Jersey	8,725	1,164 —	—	1	9 —	—	2
New York (Upstate)	11,092	1,548 —	—	—	2	1 —	—
New York City	8,214	3,262 —	—	—	—	2	1
Pennsylvania	12,441	1,750 —	—	1	11 —	—	1
E.N. Central	46,275	3,207 —	—	7	2 —	—	12
Illinois	12,832	1,348 —	—	—	1 —	—	6
Indiana	6,313	329 —	—	3 —	—	—	—
Michigan	10,096	628 —	—	—	—	—	5
Ohio	11,478	703 —	—	3	1 —	—	—
Wisconsin	5,556	199 —	—	1 —	—	—	1
W.N. Central	19,942	1,050 —	—	—	1 —	—	12
Iowa	2,982	76 —	—	—	1 —	—	—
Kansas	2,764	132 —	—	—	—	—	—
Minnesota	5,167	197 —	—	—	—	—	7
Missouri	5,843	542 —	—	—	—	—	2
Nebraska	1,768	80 —	—	—	—	—	2
North Dakota	636	8 —	—	—	—	—	1
South Dakota	782	15 —	—	—	—	—	—
S. Atlantic	57,142	10,750 —	—	1	8	2	25
Delaware	853	171 —	—	—	2 —	—	—
District of Columbia	581	871 —	—	—	—	—	—
Florida	18,090	3,961 —	—	—	1 —	—	10
Georgia	9,364	1,877 —	—	—	—	—	4
Maryland	5,616	1,394 —	—	—	2 —	—	2
North Carolina	8,856	1,024 —	—	—	1	2	6
South Carolina	4,321	742 —	—	—	1 —	—	3
Virginia	7,643	634 —	—	1 —	—	—	—
West Virginia	1,818	76 —	—	—	1 —	—	—
E.S. Central	17,755	1,693 —	—	1	2 —	—	4
Alabama	4,599	391 —	—	—	—	—	1
Kentucky	4,206	292 —	—	—	1 —	—	—
Mississippi	2,911	352 —	—	—	—	—	—
Tennessee	6,039	658 —	—	1	1 —	—	3

W.S. Central	34,186	4,303	—	3	6	—	27
Arkansas	2,811	196	—	—	2	—	1
Louisiana	4,288	879	—	—	—	—	—
Oklahoma	3,579	264	—	—	—	—	1
Texas	23,508	2,964	—	3	4	—	25
Mountain	20,845	1,517	—	5	7	—	10
Arizona	6,166	585	—	—	1	—	4
Colorado	4,753	355	—	4	2	—	2
Idaho	1,466	23	—	—	—	—	1
Montana	945	25	—	—	—	—	—
Nevada	2,495	335	—	—	N	—	2
New Mexico	1,955	113	—	1	2	—	1
Utah	2,550	68	—	—	2	—	—
Wyoming	515	13	—	—	—	—	—
Pacific	48,510	5,728	—	13	36	22	37
Alaska	670	32	—	10	—	—	—
California	36,458	4,952	—	1	35	20	33
Hawaii	1,285	78	—	—	—	—	1
Oregon	3,701	239	—	1	—	—	2
Washington	6,396	427	—	1	1	2	1
American Samoa	63	—	—	—	—	—	—
C.N.M.I.	82	—	—	—	—	—	—
Guam	171	—	—	—	—	—	—
Puerto Rico	3,928	847	—	—	—	N	—

Anthrax	Foodborne	Botulism Infant	Other§	Brucellosis	
	1	32	85	27	131
	1 —		1 —	—	
	1 —		1 —	—	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
—		2	22	3	4
—		1	9 —		2
—	—		2	1 —	
—	—	—		2	1
—		1	11 —		1
—		7	2 —		12
—	—		1 —		6

TABLE 3: Reported cases by age— United States, 2007--All 1 or 2 Case Cells

Disease	Age <1 yr	1–4 yrs	5–14 yr	15–24 yrs	25–39 yrs	40–64 yrs	>65 yrs	Age not	Total
Anthrax	—	—	1	—	—	—	—	—	1
Botulism									
foodborne	—	—	2	2	6	15	7	—	32
infant	83	2	—	—	—	—	—	—	85
other (wound & unspecified)	2	—	—	2	7	14	1	1	27
Cholera	—	—	1	2	—	3	1	—	7
Cyclosporiasis	—	—	1	—	8	25	45	14	93
California serogroup virus									
neuroinvasive	—	2	10	25	4	3	3	3	50
nonneuroinvasive	—	—	1	—	—	1	2	1	5
Eastern equine encephalitis									
neuroinvasive	—	1	—	1	—	—	—	1	3
nonneuroinvasive	—	—	—	—	—	1	—	—	1
Powassan virus disease, neuroinvasive	—	—	1	—	—	—	2	4	7
St. Louis encephalitis virus disease									
neuroinvasive	—	—	1	1	—	—	3	3	8
nonneuroinvasive	—	—	—	—	1	—	—	—	1
West Nile virus disease									
neuroinvasive	—	1	5	29	58	138	532	463	1,227
Ehrlichiosis									
human monocytic	—	2	6	50	61	102	383	219	828
human (other and unspecified)	—	1	13	28	13	35	154	85	337
Hantavirus pulmonary syndrome	—	—	—	1	8	11	11	1	32
Hemolytic uremic syndrome	—	6	138	97	16	2	14	14	292
Hepatitis, viral, acute									
B	—	6	1	8	370	1,783	2,062	249	4,519
C	—	8	—	1	133	319	352	26	845
Measles, total	—	7	4	3	13	7	7	1	43
Meningococcal disease									
all serogroup	—	68	66	37	123	47	64	64	492
serogroup A, C, Y, & W-13	—	28	31	27	87	35	60	56	325
serogroup b	—	40	35	10	36	12	24	8	167
Novel influenza A virus infection	—	—	1	1	—	1	1	—	4
Plague	—	—	1	—	—	1	5	—	7
Psittacosis	—	—	—	1	1	5	4	1	12
Q fever	—	—	1	2	9	36	88	34	171
Rabies, human	—	—	—	—	—	—	1	—	1
Rocky Mountain spotted fever	—	2	41	178	221	452	1,022	299	2,221
Rubella	—	—	1	—	5	6	—	—	12
Streptococcal toxic-shock syndrome	—	—	6	1	9	20	54	42	132
Syphilis, primary and secondary	—	1	1	13	2,481	5,059	3,816	91	11,466
Tetanus	—	—	—	2	2	4	5	10	28
Toxic-shock syndrome	—	—	6	21	22	12	30	1	92

Trichinellosis	—	—	—	—	1	3	1	—	5
Tuberculosis***	115	351	313	1,581	3,266	5,093	2,578	2	13,299
Tularemia	—	18	27	11	16	44	20	1	137
Vancomycin-intermediate S	—	—	—	1	2	20	14	—	37
Vancomycin-resistant Staph	—	—	—	—	—	2	—	—	2
Vibriosis	3	8	58	47	94	225	113	1	549
Totals	308	683	903	5,204	11,462	14,099	4,361	90	37,110

TABLE 3: Reported cases by age— United States, 2007--All 1 or 2 Case Cells

Disease	<1 yr	1–4 yrs	5–14 yrs	15–24 yrs	25–39 yrs	40–64 yrs	>65 yrs	Age not sta	Total
Anthrax	—	—	1	—	—	—	—	—	1
Botulism									
foodborn	—	—	2	2	6	15	7	—	32
infant	83	2	—	—	—	—	—	—	85
other (w	2	—	—	2	7	14	1	1	27
Cholera	—	—	1	2	—	3	1	—	7
Cyclospori	—	1	—	8	25	45	14	—	93
California serogroup virus									
neuroinv	2	10	25	4	3	3	3	—	50
nonneur	—	1	—	—	1	2	1	—	5
Eastern eq									
neuroinv	1	—	1	—	—	—	1	—	3
nonneur	—	—	—	—	1	—	—	—	1

TABLE 3: Reported cases by age— United States, 2007--All 1 or 2 Case Cells

Disease	<1 yr	1–4 yrs	5–14 yrs	15–24 yrs	25–39 yrs	40–64 yrs	>65 yrs	Age not sta	Total
Anthrax	—	—	2	—	—	—	—	—	2
Botulism									
foodborn	—	—	2	2	6	15	7	—	32
infant	83	2	—	—	—	—	—	—	85
other (w	2	—	—	2	7	14	0	2	27
Cholera	—	—	0	2	—	2	2	—	6
Cyclospori	—	2	—	8	24	45	14	—	93
California serogroup virus									
neuroinv	2	10	25	4	3	3	3	—	50
nonneur	—	0	—	—	0	2	2	—	4
Eastern eq									
neuroinv	2	—	0	—	—	—	2	—	4
nonneur	—	—	—	—	2	—	—	—	2

REPORTED CASES

State = ??? Year

= 200?

Original Data

Diagnosis	Age group										Total
	<1	1 to 4	5 to 9	10 to 14	15-19	20-39	40-59	60-79	80+	Unk	
AMEBIASIS
ANIMAL BITES	4	96	107	98	47	187	186	83	13	24	845
ANTHRAX
BOTULISM, FOODBORNE	2	1	1	.	.	4
BOTULISM, INFANT	2	2
BOTULISM, OTHER
BRUCELLOSIS	.	1	1	.	.	.	2
CAMPYLOBACTER	13	58	42	24	49	248	272	95	15	.	816
CARBON MONOXIDE POISONING
CHOLERA
CRYPTOSPORIDIOSIS	3	35	31	22	14	77	18	13	3	.	216
CYCLOSPORIASIS	1	1
DENGUE FEVER
DIPHThERIA
ENCEPHALITIS OTHER	.	2	5	1	.	5	4	5	1	.	23
GIARDIASIS	5	102	65	19	37	149	153	43	7	.	580
GROUP A STREP INVASIVE	4	6	5	2	3	31	40	40	12	1	144
GROUP B STREP INVASIVE	16	.	.	.	2	19	44	32	15	.	128
HAEMOPHILUS INFLUENZAE	7	3	1	1	1	7	14	15	9	.	58
HANTAVIRUS PULMONARY SYNDROME	4	2	1	.	.	7
HEMOLYTIC UREMIC SYNDRM	.	2	1	.	1	4
HEPATITIS A	.	2	.	1	2	9	9	2	2	.	27
HEPATITIS B, ACUTE	23	18	.	.	.	41
HEPATITIS B, CHRONIC	.	7	8	6	14	260	167	62	3	.	527
HEPATITIS B, PERINATAL INFECTION	.	2	2
HEPATITIS C, ACUTE	11	10	.	.	.	21
HEPATITIS C, CHRONIC	10	1	1	4	24	840	2305	243	26	9	3463
HEPATITIS D	1	1	.	1	.	.	3
INFLUENZA-hospitalized	68	70	32	4	8	50	78	77	56	.	443
INFLUENZA-pediatric death	.	.	1	1
KAWASAKI SYNDROME	5	29	5	3	42
LEGIONELLOSIS	3	14	5	.	.	22
LEPROSY (HANSEN DIS)	1	.	.	.	1
LISTERIOSIS	1	2	6	2	.	11
LYME DISEASE
MALARIA	.	.	2	1	3	12	5	3	.	.	26

MEASLES
MENINGITIS ASEPTIC/VIRAL	81	13	20	27	30	144	81	21	7	.	424
MENINGITIS OTH BACT
MENINGOCOCCAL DISEASE	1	3	.	.	4	3	4	4	3	.	22
MUMPS	.	3	6	.	1	5	1	1	.	.	17
PERTUSSIS	38	42	29	36	36	51	50	20	4	.	306
PLAGUE
POLIO
PSITTACOSIS	1	1	.	.	.	2
RABIES ANIMAL
RABIES HUMAN
RELAPSING FEVER	1	1
ROCKY MTN SPOTTED FVR	.	.	1	.	.	2	3
RUBELLA
RUBELLA CONGENITAL
SALMONELLOSIS	47	58	45	42	30	132	135	62	12	.	563
SHIGELLOSIS	2	22	20	4	5	31	28	6	4	.	122
STEC (shiga toxin producing E.coli)	4	31	17	10	13	48	20	8	3	.	154
STREP PNEUMO INVASIVE	22	29	12	5	7	46	155	120	63	.	459
TETANUS
TOXIC SHOCK-OTHER	.	2	.	.	1	.	1	.	.	.	4
TOXIC SHOCK-STREP	1	.	.	1
TRICHINOSIS
TULAREMIA	.	2	1	3
TYPHOID FEVER	.	1	.	1	1	3	6
VACCINIA, CONTACT TRANSMISSION
VARICELLA(CHICKEN POX)	24	126	505	336	45	37	13	3	.	.	1089
WEST NILE VIRUS	.	1	3	8	26	130	282	112	16	.	578
YERSINIOSIS	.	.	1	.	1	1	.	1	.	.	4
Total	356	749	966	655	406	2575	4115	1086	276	34	11218

REPORTED CASES BY AGE: PRIMARY SUPPRESSION

THRESHOLD = 3; Partial Table

4	96	107	98	47	187	186	83	13	24	845
0	0	0	0	0	*	*	*	0	0	4
*	0	0	0	0	0	0	0	0	0	*
0	*	0	0	0	0	*	0	0	0	*
13	58	42	24	49	248	272	95	15	0	816
0	0	0	0	0	*	0	0	0	0	*
0	*	*	0	*	0	0	0	0	0	4
0	*	0	0	0	0	0	0	0	0	*
10	*	*	4	24	840	2305	243	26	9	3463
0	0	0	0		*	0	*	0	0	3
68	70	32	4	8	50	78	77	56	0	443
0	0	*	0	0	0	0	0	0	0	*
0	0	0	0	0	0	*	0	0	0	*
38	42	29	36	36	51	50	20	4	0	306
0	0	0	0	0	*	*	0	0	0	*
0	0	0	0	0	*	0	0	0	0	*
0	0	*	0	0	*	0	0	0	0	3
4	31	17	10	13	48	20	8	3	0	154
22	29	12	5	7	46	155	120	63	0	459
0	*	0	0	*	0	*	0	0	0	4
0	0	0	0	0	0	0	*	0	0	*
0		*	0	0	0	0	0	0	0	3
24	126	505	336	45	37	13	3	0	0	1089
0	0	*	0	*	*	0	*	0	0	4
185	462	750	517	233	1516	3084	653	180	33	7613

* = 1 or 2 cases

REPORTED CASES BY AGE
OPTIMAL COMPLEMENTARY SUPPRESSION
THRESHOLD = 3

4	96	107	98	47	187	186	83	13	24	845
0	0	0	0	0	*	*	*	0	0	4
*	0	0	0	0	0	0	0	0	0	*
0	*	0	0	0	0	*	0	0	0	*
13	58	42	24	49	248	272	95	15	0	816
0	0	0	0	0		0	0	0	0	*
0	*	*	0	*	0	0	0	0	0	4
0	*	*	0	0	0	0	0	0	0	*
*	*	*	4	24	840	2305	243	26	9	3463
0	0	0	0	*	*	0	*	0	0	3
68	70	32	4	8	50	78	77	56	0	443
0	0	*	0	0	0	0	0	0	0	*
0	0	0	0	0	0	*	0	0	0	*
38	42	29	36	36	51	50	20	4	0	306
0	0	0	0	0	*	*	0	0	0	*
0	0	0	0	0	*	0	0	0	0	*
0	0	*	0	0	*	0	0	0	0	3
4	31	17	10	13	48	20	8	3	0	154
22	29	12	5	7	46	155	120	63	0	459
0	*	0	0	*	0	*	0	0	0	4
0	0	0	0	0	0	0	*	0	0	*
0	*	*	0	0	0	0	0	0	0	3
24	126	505	336	45	37	13	3	0	0	1089
0	0	*	0	*	*	0	*	0	0	*
185	462	750	517	233	1516	3084	653	180	33	7613

* = suppression

Deaths by Cause and Race-Ethnicity
County = ???, Year = 200?
Released Table: Threshold = 3

	All Races	White Non-Hispan	White Hispan	Black	Asian Amer/ Pacific Island	Amer Indian/ Nativ Alask
Total	1366	1130	220	*	*	6
Cardiovascular Disease	365	317	*	*	*	0
Heart Disease	287	250	*	*	0	0
Cerebrovas Disease	61	51	*	0	*	0
Atherosclerosis	*	*	*	0	0	0
Malignant Neoplasms	301	249	47	*	*	*
Lung Cancer	72	67	*	0	0	*
Breast Cancer	31	23	8	0	0	0
Chron Low Resp Disease	104	95	9	0	0	0
Unintentional Injuries	97	76	*	0	*	0
Motor Vehicle	35	26	9	0	0	0
Other Unintent Injur	62	50	10	0	*	0
Pneumonia & Influenza	24	21	*	0	0	0
Suicide	28	22	6	0	0	0
Diabetes Mellitus	56	39	17	0	0	0
HIV Infection	*	*	*	0	0	0
Chron Liver/Cirrhosis	27	17	*	0	0	*
Alzheimer's Disease	74	66	*	0	0	*
Nephritis, Nephrosis,	10	*	*	0	0	0
Homicide/Legal Interv	7	0	7	0	0	0
Septicemia	13	10	*	0	*	0
Congenital Anomalies	*	4	*	0	0	0
Perinatal Period Cond	13	6	7	0	0	0
All Other	239	198	38	*	0	*

*** = 1 or 2 cases**

County Deaths by Cause and Race-Ethnicity
Complementary Suppression: Threshold = 3

W-NH	W-H	Blk	AA/PI	AI/NA	Unk	
317	**	*	*	0	0	365
250	**	*	0	0	0	287
51	**	0	*	0	0	61
*	*	0	0	0	0	*
249	47	*	*	*	*	301
67	**	0	0	*	0	72
23	8	0	0	0	0	31
95	9	0	0	0	0	104
76	**	0	*	0	0	97
26	9	0	0	0	0	35
50	10	0	*	0	0	62
21	*	0	0	0	*	24
22	6	0	0	0	0	28
39	17	0	0	0	0	56
*	*	0	0	0	0	*
17	**	0	0	*	0	27
66	**	0	0	*	0	74
**	*	0	0	0	0	10
0	7	0	0	0	0	7
10	*	0	*	0	0	13
4	*	0	0	0	0	**
6	7	0	0	0	0	13
198	38	*	0	*	0	239
1130	220	**	**	6	*	1366

* = 1 or 2 cases

** = complementary suppression (for comprehension only)

QUESTION: Is 7 cases all in one category disclosure?

County Deaths by Cause and Race-Ethnicity
Fully Reconstructed Table: Subcategories Removed
Threshold = 3

W-NH	W-H	Blk	AA/PI	AI/NA	Unk	
317	46	1	1	0	0	365
249	47	1	1	2	1	301
95	9	0	0	0	0	104
76	19	0	2	0	0	97
21	2	0	0	0	1	24
22	6	0	0	0	0	28
39	17	0	0	0	0	56
1	1	0	0	0	0	2
17	9	0	0	1	0	27
66	7	0	0	1	0	74
9	1	0	0	0	0	10
0	7	0	0	0	0	7
10	2	0	1	0	0	13
4	2	0	0	0	0	6
6	7	0	0	0	0	13
198	38	1	0	2	0	239
1130	220	3	5	6	2	1366

County Deaths by Cause and Race-Ethnicity

CTA Treated: Subcategories Removed

Threshold = 2

W-NH	W-H	Blk	AA/PI	AI/NA	Unk	
317	46	0	2	0	0	365
249	48	2	0	2	0	301
95	9	0	0	0	0	104
76	19	0	2	0	0	97
22	0	0	0	0	2	24
22	6	0	0	0	0	28
39	17	0	0	0	0	56
0	2	0	0	0	0	2
16	9	0	0	2	0	27
65	7	0	0	2	0	74
10	0	0	0	0	0	10
0	7	0	0	0	0	7
11	2	0	0	0	0	13
4	2	0	0	0	0	6
6	7	0	0	0	0	13
198	39	2	0	0	0	239
1130	220	4	4	6	2	1366

County Deaths by Cause and Race-Ethnicity
Fully Reconstructed Table: Subcategories Removed
Threshold = 3

W-NH	W-H	Blk	AA/PI	AI/NA	Unk	
317	45	3	0	0	0	365
249	47	0	0	3	0	299
95	9	0	0	0	0	104
76	18	0	3	0	0	97
21	0	0	0	0	3	24
22	6	0	0	0	0	28
39	17	0	0	0	0	56
0	3	0	0	0	0	3
17	10	0	0	0	0	27
66	8	0	0	0	0	74
9	1	0	0	0	0	10
0	7	0	0	0	0	7
11	2	0	0	0	0	13
4	2	0	0	0	0	6
6	7	0	0	0	0	13
198	38	0	0	3	0	239
1130	220	3	3	6	3	1365

**LIVE BIRTHS BY EDUCATION, AGE OF MOTHER,
AREA WITHIN STATE = 3, RACE = White**

EDUCATION		AGE		
		Under 15 Years	15	16
Total	2,171	1	3	3
Under 9 Years	29	-	1	-
9-11 Years	13	-	1	2
12 Years	102	-	-	1
13-15 Years	180	-	-	-
16+ Years	1,798	1 ERROR!	.	
Not Stated	49	-	1	-
Area 3	841	1	1	-

**So: 1 White, Under 15 Years Mother in Area 3
--Education Data (16+) Erroneous**

**LIVE BIRTHS BY EDUCATION, AGE OF MOTHER,
AREA = 3, ETHNICITY = Non-Hispanic**

Education	Total	Age Of Mother		
		Under 15	15	16
Area 3	907	1	-	1
16+ Years	795	1	-	-

So: 1 White, Under 15 Years, Non-Hispanic Mother in Area 3

LIVE BIRTHS BY MARITAL, HISPANIC, AGE OF MOTHER, AREA

Marital Status, Hispanic: Area 3

	Total	Non-Hispanic	Hispani	Married Total
Area 3	936	907	29	890
Under 15 Years	1	1	-	1

So: 1 White, Under 15 Years, Non-Hispanic, Married Mother in Area 3

LIVE BIRTHS BY RACE, PRIOR LIVE BIRTH, AGE OF MOTHER

		Under 15 Years	15	16
White	2,171	1	-	3
None	1,300	1	-	3
One	646	-	-	-

So: 1 White, Under 15 Years, Non-Hispanic, Married Mother with No Previous Live Births in Area 3

LIVE BIRTHS BY RACE, BIRTH WEIGHT, AGE OF MOTHER

		Under 15 Years	15
White	2,171	1	-
3500-4499 grams	869	1	-
4500+ grams	32	-	-

So: 1 White, Under 15 Year, Non-Hispanic, Married Mother with No Previous Live Births and Birth Weight 3500-4499 grams in Area 3

LIVE BIRTHS BY BIRTH WEIGHT, WEEKS GESTATION, RACE, AGE OF MOTHER

		Under 15	15
2,500 grams Or Mc	7,052	19	47
White	2,021	1	-
Under 36 Weeks	81	-	-
36-37 Weeks	148	-	-
38-39 Weeks	1,025	1	-

So: 1 White, Under 15 Year, Non-Hispanic, Married Mother with No Previous Live Births, Birth Weight 3500-4499 grams and 38-39 Week Gestation in Area 3

**LIVE BIRTHS BY RACE, AGE OF MOTHER,
TRIMESTER PRENATAL CARE BEGAN**

		Under 15	15
Total	7,940	23	52
White	2,171	1	-
First Trimester	1,972	-	-
Second Trimester	147	1	-
Third Trimester	23	-	-
No Prenatal Care	4	-	-
Not Stated	25	-	-

**So: 1 White, Under 15 Year, Non-Hispanic, Married Mother
with No Previous Live Births, Birth Weight
3500-4499 grams 38-39 Week Gestation,
Began Prenatal Care in Trimester 2 in Area 3**

ADDITIONAL EXAMPLES (State Level/Release)

- 1 Under 15, Other Race, Hispanic Unmarried Mother
With Less Than 9 Years Education in Area 4
- 1 Age 15, White, Non-Hispanic Married Mother
With 9-11 Years Education in Area 3
- 6 Age 15, Black, Non-Hispanic Unmarried Mothers
With Various Education Levels in Area 2
- 2 Under 15, Black, Hispanic Unmarried Mothers
With 9-12 Years Education in Area 1
- 3 Age 15, White Mothers:
 - 1 Non-Hispanic, Less Than 9 Years Education in Area 2
 - 1 Non-Hispanic, 9-11 Years Education in Area 3
 - 1 Hispanic, Indeterminate Education in Area 4
- 1 Age 18, Black, Non-Hispanic With 4+ Prior Live Births
- 1 Under 15 Mother With Birth Less than 500 Grams in Area 5
- 1 Under 15 Mother With Birth 1500-1999 Grams in Area 7

Area Populations Clustered Around 70,000

Transparency in SDL

- *Kerckhoff's principle* of cryptography: a cryptosystem should be secure even if everything about the system, except the keys, is known to the intruder
- Statistical agencies adopted conservative SDL policies-- very little about the masking is made known to the user
 - distribution of random errors (perturbation)
 - swap rate or strategy (swapping/switching)
 - models (synthesizing)
 - tight bounds on suppressed entries (suppression)
- Agencies believe this is necessary to or enhances protection
- Some work has shown that this is in fact the case
- But information on injected errors/biases is important to data users and interpretation of inferences
- Serious work is needed on Transparency in SDL

Concluding Comments

- Public health data should, and are required to be, protected
- Rules addressing risk of disclosure are variable
 - Between public health agencies
 - Within public health agencies
- Methods employed to protect public health data
 - Typically substandard
 - Often ineffective
 - Should be replaced by statistical methods
- This, however, is not always easy
 - Zero counts, less important in social science, often are very important in public health
 - Tables can be many-dimensional
 - Data may appear in many places (US, state, etc.) and forms (weekly/monthly/annual reports)
 - The quality of disclosure-limited data needs to be preserved
- Transparency should be studied
- Beware of software providers bearing gifts
 - “I have your solution”
 - First, they must understand the problem!