To Suppress or Not to Suppress…that is the question

RI BRFSS

2017 Northeast Epidemiology Conference
October 19, 2017

Objectives

1) Present an overview of the implementation of RIDOH policy for dealing with small numbers using two examples from RI BRFSS

   - Identifying "small numbers"
   - Employing strategies to avoid suppression

2) Discuss common practices on data collection, analysis and dissemination related to small numbers and data reliability.

Why we care about small numbers?

Public health policy decisions are informed by statistical information.

- Small numbers are associated with estimates of lower statistical credibility and tend to be unstable and prone to misinterpretation, which could mislead policy decisions.
- Small numbers may facilitate identification of individuals.

As epidemiologists, we make decisions whether data can be reported without reservation, with caution, or suppressed entirely.

Rules for suppression, Reporting of BRFSS results

- Numerator less than 10?

- Denominator less than 50?

- Confidence interval more than 20 points wide?

- Use of half width confidence intervals?

- Relative standard error (RSE ≥30%)

$$RSE = \frac{\text{Standard error of estimate}}{\text{Estimate}} * 100$$

- Unreliable estimates, caution note?

- SUPPRESSION as last resort

Dealing with small numbers in RI

RIDOH Epidemiologists and Evaluators developed a small numbers policy in 2015

- Confidence intervals (CIs) do not represent a reliable standard when numerators, denominators, or both are small.

- Recommends Relative Standard Error (RSE) as the gold standard for determining whether data be suppressed.

- Criteria
  - Estimates with RSE ≥30% are suppressed
  - If RSE is 20 - <30%, estimate is unstable and needs to be interpreted with caution

Example 1

Collapsing strata

Initial Output

Numerators < 20!   Now what?   Calculate RSEs

	Straight (n=4,597, 94.7%)		Gay/Lesbian/Bisexual/Other (n=219; 5.3%)	
	N	%	N	%
Sex				
Male	1,874	48.0	101	46.0
Female	2,723	52.0	118	54.0
Age Group				
18-29 yrs	313	18.2	41	38.7
30-44 yrs	677	24.2	37	23.9
45-64 yrs	1,945	35.3	98	28.5
65+ yrs	1,587	22.3	43	8.9
Race/ethnicity				
White only, Non-Hispanic	3,888	77.9	165	67.1
Black only, Non-Hispanic	152	4.8	8	4.1
Other race/Multiracial,Non-Hispanic	205	6.2	12	9.0
Hispanic	292	11.0	31	19.7

Source: 2016 BRFSS

Calculating RSEs

Goal is not to suppress, so we will collapse race/ethnicity categories.

Sexual Orientation and Other Demographics, 2016

	Straight (n=4,597, 94.7%)			Gay/Lesbian/ Bisexual/Other (n=219; 5.3%)		
	N	%	RSE%	N	%	RSE%
Sex						
Male	1,874	48.0	1.83	101	46.0	9.94
Female	2,723	52.0	1.21	118	54.0	9.18
Age Group						
18-29 yrs	313	18.2	5.52	41	38.7	15.55
30-44 yrs	677	24.2	3.56	37	23.9	16.38
45-64 yrs	1,945	35.3	1.74	98	28.5	10.00
65+ yrs	1,587	22.3	2.05	43	8.9	15.18
Race/ethnicity						
White only, Non-Hispanic	3,888	77.9	0.69	165	67.1	7.65
Black only, Non-Hispanic	152	4.8	7.98	8	4.1	35.33
Other race/Multiracial,Non-Hispanic	205	6.2	6.83	12	9.0	28.83
Hispanic	292	11.0	5.67	31	19.7	17.90

Source: 2016 RI BRFSS

Collapsing strata and calculating RSEs

RSE 20-30%   Add caution note

Sexual Orientation and Other Demographics, 2016 - Race/Ethnicity Strata Collapsed

	Straight (n=4,597, 94.7%)			Gay/Lesbian/Bisexual/Other (n=219; 5.3%)		
	N	**%**	**RSE%**	**N**	**%**	**RSE%**
Race/ethnicity						
White only, Non-Hispanic	3,888	77.9	0.69	165	67.1	7.65
Other race/Multiracial,Non-Hispanic	357	11.0	5.10	20	13.2	22.3
Hispanic	292	11.0	5.67	31	19.7	17.90

Source: 2016 RI BRFSS

Final display of data

Sexual Orientation and Other Demographics, 2016

	Straight (n=4,597, 94.7%)			Gay/Lesbian/ Bisexual/Other (n=219; 5.3%)		
	N	**%**	**95% CI**	**N**	**%**	**95% CI**
Sex						
Male	1,874	48.0	[45.9-50.1]	101	46.0	[36.2-55.8]
Female	2,723	52.0	[49.9-54.1]	118	54.0	[44.2-63.8]
Age Group						
18-29 yrs	313	18.2	[16.0-20.4]	41	38.7	[28.4-49.0]
30-44 yrs	677	24.2	[22.2-26.2]	37	23.9	[15.1-32.7]
45-64 yrs	1,945	35.3	[33.4-37.1]	98	28.5	[21.2-35.7]
65+ yrs	1,587	22.3	[20.9-23.6]	43	8.9	[5.4-12.5]
Race/ethnicity						
White only, Non-Hispanic	3,888	77.9	[76.0-79.9]	165	67.1	[57.6-76.6]
Other race/Multiracial,Non-Hispanic	357	11.0	[9.5-12.5]	20	13.2*	[5.6-20.7]
Hispanic	292	11.0	[9.5-12.6]	31	19.7	[12.1-27.3]

Source: 2016 RI BRFSS

Note: * RSE>20%; Estimate should be interpreted with caution.

Example 2

Aggregating multiple years of data

Initial Output

Numerators < 20!   Now what?   Calculate RSEs

Myocardial Infarction Demographics, 2016

	No Mycardial Infarction N=5116; 95.5%			Mycardial Infarction N=314; 4.5%		
	N	**%**	**95% CI**	**N**	**%**	**95% CI**
Sex						
Male	2,048	9.9	[92.7-95.1]	176	6.1	[4.9-7.3]
Female	3,068	97.0	[96.4-97.7]	138	3.0	[2.3-3.6]
Age Group						
18-44 yrs	1,231	98.8	[98.0-99.6]	14	1.2	[0.4-2.0]
45-64 yrs	2,154	94.9	[93.7-96.1]	113	5.1	[3.9-6.3]
65+ yrs	1,632	89.6	[87.8-91.3]	181	10.4	[8.7-12.2]
Race/ethnicity						
White only, Non-Hispanic	4,227	95.5	[94.7-96.2]	256	4.5	[3.8-5.3]
Black only, Non-Hispanic	175	97.2	[95.1-99.2]	10	2.8	[0.7-4.9]
Other race/Multiracial,Non-Hispanic	235	95.3	[92.1-98.5]	15	4.7	[1.5-7.9]
Hispanic	392	95.7	[93.4-98.0]	24	4.3	[2.0-6.6]

Source: 2016 RI BRFSS

Calculating RSEs

Goal is not to suppress, so we will aggregate data with 2015 to achieve a more stable estimate for subgroups.

Myocardial Infarction Demographics, 2016

	No Mycardial Infarction N=5116; 95.5%			Mycardial Infarction N=314; 4.5%		
	N	%	RSE%	N	%	RSE%
Sex						
Male	2,048	9.9	1.74	176	6.1	7.41
Female	3,068	97.0	1.19	138	3.0	8.40
Age Group						
18-44 yrs	1,231	98.8	2.50	14	1.2	26.69
45-64 yrs	2,154	94.9	1.66	113	5.1	9.31
65+ yrs	1,632	89.6	2.06	181	10.4	7.31
Race/ethnicity						
White only, Non-Hispanic	4,227	95.5	0.70	256	4.5	6.10
Black only, Non-Hispanic	175	97.2	7.43	10	2.8	31.59
Other race/Multiracial,Non-Hispanic	235	95.3	6.38	15	4.7	25.78
Hispanic	392	95.7	4.86	24	4.3	20.37

Source: 2016 RI BRFSS

Aggregating two years of data and calculating RSEs

Check RSEs,   <20%

Myocardial Infarction Demographics, 2015 and 2016 combined

	No Mycardial Infarction N=10,924; 95.8%			Mycardial Infarction N=670; 4.2%		
	N	%	RSE%	N	%	RSE%
Sex						
Male	4,360	94.0	1.20	397	6.0	4.93
Female	6,564	97.4	0.81	273	2.6	5.98
Age Group						
18-44 yrs	2,574	99.1	1.73	25	0.9	19.98
45-64 yrs	4,491	95.3	1.16	221	4.7	6.66
65+ yrs	3,678	89.4	1.36	415	10.6	4.82
Race/ethnicity						
White only, Non-Hispanic	9,051	95.6	0.48	556	4.4	4.14
Black only, Non-Hispanic	379	97.1	5.05	25	2.9	19.98
Other race/Multiracial,Non-Hispanic	505	96.7	4.35	29	3.3	18.55
Hispanic	793	95.9	3.43	41	4.1	15.59

Source: 2015-2016 RI BRFSS

Final display of data

Myocardial Infarction Demographics, 2015 and 2016 combined

	No Mycardial Infarction N=10,924; 95.8%			Mycardial Infarction N=670; 4.2%		
	N	**%**	**95% CI**	**N**	**%**	**95% CI**
Sex						
Male	4,360	94.0	[93.2-94.8]	397	6.0	[5.2-6.8]
Female	6,564	97.4	[96.9-97.8]	273	2.6	[2.2-3.1]
Age Group						
18-44 yrs	2,574	99.1	[98.6-99.6]	25	0.9	[0.4-1.4]
45-64 yrs	4,491	95.3	[94.5-96.1]	221	4.7	[3.9-5.5]
65+ yrs	3,678	89.4	[88.2-90.6]	415	10.6	[9.4-11.8]
Race/ethnicity						
White only, Non-Hispanic	9,051	95.6	[95.1-96.1]	556	4.4	[3.9-4.9]
Black only, Non-Hispanic	379	97.1	[95.7-98.5]	25	2.9	[1.5-4.3]
Other race/Multiracial,Non-Hispanic	505	96.7	[95.0-98.5]	29	3.3	[1.5-5.0]
Hispanic	793	95.9	[94.1-97.6]	41	4.1	[2.4-5.9]

Source: 2015-2016 RI BRFSS

Input from audience regarding issues related to small numbers

- Best way to calculate RSE?
  - SAS macro available?
- Policy for adding new BRFSS items when behavior/condition has low prevalence
  - Committing to multiple consecutive years of collection?
- How to address high item non-response?
  - Placement of SAQs and drop-off
- Raw data file provided to partners?
  - Technical assistance to avoid misuse of data

Acknowledgments

Alvaro Tinajero, MD MPH ScM
alvaro.tinajero@health.ri.gov
RIDOH Statistical Reliability and Confidentiality Work Group

RHODE ISLAND DEPARTMENT OF HEALTH

Tara Cooper MPH
RI BRFSS/YRBS Coordinator
401-222-7628
tara.cooper@health.ri.gov

Tracy Jackson PhD
Sr. Public Health Epidemiologist
401-222-7465
tracy.jackson@health.ri.gov