Making the Most of Your Surveillance Data:

Biostatistics for Infection Control

Emily Sickbert-Bennett, PhD, MS, CIC, FSHEA

Director of Infection Prevention, UNC Medical Center
Professor of Medicine-Infectious Diseases,
UNC School of Medicine
Associate Professor of Epidemiology,
Gillings School of Global Public Health

1

Statistics

The margin of error...

17 in every 100 people...

Men are at 3 times higher risk...

Numbers that describe the health of the population

1 in 9 children...

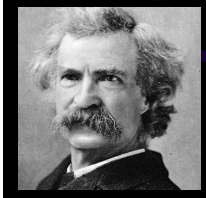
39% OF THE POPULATION...

The **science** used to interpret these numbers.

There is a statistically significant difference...

Risk of dying is 8 times higher among...

2



"There are 3 kinds of lies. Lies, damned lies, and statistics."

~Popularized by Mark Twain

- Describes the persuasive power of numbers, particularly the use of statistics, to bolster weak arguments, and the tendency of people to disparage statistics that do not support their positions.

3

Outline

- Describe Surveillance Data
- Display and Interpret Surveillance Data
- Determine the Significance of Changes to Surveillance Data

4

Describing Surveillance Data Using Descriptive Statistics

5

Absolute Measures

- Simplest type of measurement
- Also known as counts
- Example:
  - Hospital A: 25 patients with norovirus
  - Hospital B: 10 patients with norovirus
- Is norovirus worse at Hospital A?

6

Relative Measures

- Includes a denominator
- Useful for comparisons
- Examples:
  - 16 cases of *C. difficile* out of 1000 patients
  - 1 positive *C. difficile* test out of 7 samples tested

7

Absolute versus Relative

Example: Norovirus activity at different hospitals

- Absolute measures
  - Hospital A: 25 patients ill
  - Hospital B: 10 patients ill
- Relative measures
  - Hospital A: 25 ill per 1000 patients = 0.025 or 2.5%
  - Hospital B: 10 ill per 250 patients = 0.040 or 4%

8

Descriptive Statistics

- Measures of Rates and Ratios
  - *Rate: How fast disease occurs in a population.*
  - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
  - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion (Variability)
  - *Dispersion: How widely your data is spread from the average.*

9

What Makes a Rate?

1. Numerator (top number)
   - *e.g., number of infections*
2. Denominator (bottom number)
   - *e.g., number of patients [proportion]*
   - *e.g., number of patient-days, number of device-days [incidence density/rate]*
3. Time Frame
   - *e.g., day, week, month*

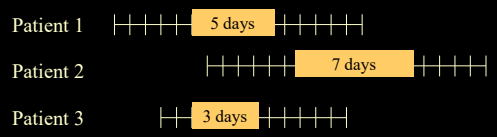
10

Denominators

- Represent the population at risk of becoming part of the numerator
- Often, the most difficult data to obtain, but essential for comparisons
- Ideally, should incorporate time and can account for risk factors such as device use (e.g., device-days), length of stay (e.g., patient-days)

11

What is a Patient/Device-Day?



=15 patient-days, device-days, etc.

- Gives more information than simply—3 patients
- Strategies: e.g., count how many at 9 am

12

Rate Measures

- Prevalence
- Incidence
- Attack Rate

13

Prevalence

- Prevalence: the total number of cases of disease existing in a population at a point in time.
  - *e.g., # of MRSA cases per population **on** March 8*

Count of existing cases x constant (e.g., 100 or 1000) = Number of people at risk

14

Incidence

- Incidence: the number of new cases of disease in a population over a period of time.
  - *e.g., # of **new** MRSA cases per population **during** March*

Count of new cases x constant (e.g., 100 or 1000) = Number of people at risk

15

Attack Rate

- Attack Rate: the number of new cases of disease out of the population at risk.
  - Related to incidence but always uses 100 as the constant, so it is expressed as a percent.
  - Often used for outbreaks or clusters that occur over a short period of time
  - *e.g., **%** of patients with MRSA during outbreak in Med ICU in March*

Count of new cases x 100 = Number of people at risk

16

Example 1:

- You perform HAI surveillance for ventilator associated pneumonias (VAP) and central line associated bloodstream infections (CLABSI) in your 12 bed intensive care unit.
- In March, you identify 2 new VAPs, 4 new CLABSIs and 3 new respiratory infections (not ventilator associated).

17

Example 1:

- The admitting department tells you that in March there were 89 patients in the unit with 311 patient-days.
- Respiratory care tells you that they provided 162 ventilator-days of care to 47 patients in March.
- You count the central line-days and find 284 line-days in 84 patients in March.

18

Example 1:

- In March, what was the VAP rate?
  - Incidence or prevalence?
  - Numerator?
  - Denominator?
  - Units?

19

Example 1: Answers

- In March, what was the VAP rate?
  - Incidence or prevalence?
    - Incidence
  - Numerator?
    - 2
  - Denominator?
    - 162 or 47
  - Units?
    - "infections per 1000 ventilator-days" or "infections per 100 ventilated patients during March"
  - ANSWER: 12.3 infections per 1000 ventilator-days; 4.3 infections per 100 ventilated patients during March.

20

Example 1:

- In March, what was the CLA-BSI rate?
  - Incidence or prevalence?
  - Numerator?
  - Denominator?
  - Units?

21

Example 1: Answers

- In March, what was the CLA-BSI rate?
  - Incidence or prevalence?
    - Incidence
  - Numerator?
    - 4
  - Denominator?
    - 284 or 84
  - Units?
    - "infections per 1000 central line-days" or "infections per 100 patients with central lines during March"
  - ANSWER: 14.1 infections per 1000 central line-days or 4.8 infections per 100 patients with central lines during March

22

Example 1:

- In March, what was overall infection rate?
  - Incidence or prevalence?
  - Numerator?
  - Denominator?
  - Units?

23

Example 1: Answers

- In March, what was overall infection rate?
  - Incidence or prevalence?
    - Incidence
  - Numerator?
    - 9
  - Denominator?
    - 311 or 89
  - Units?
    - "infections per 1000 patient-days" or "infections per 100 patients during March"
  - ANSWER: 28.9 infections per 1000 patient-days or 10.1 infections per 100 patients during March

24

Example 1:

- On April 7, you were worried about the BSI rate so you return to the unit to do a "spot check" on all of the patients for a BSI.
- At that time with a census of 12, you reviewed 11 charts and found 1 healthcare associated BSI.

25

Example 1:

- On April 7th, what was the BSI infection rate at the time of your spot check?
  - Incidence or prevalence?
  - Numerator?
  - Denominator?
  - Units?

26

Example 1: Answers

- In April, what was the BSI infection rate at the time of your spot check?
  - Incidence or prevalence?
    - Prevalence
  - Numerator?
    - 1
  - Denominator?
    - 11
  - Units?
    - "prevalent infections per 100 patients on April 7th"
  - ANSWER: 9 prevalent infections per 100 patients on April 7th.

27

What Makes a Standardized Infection Ratio (SIR)?

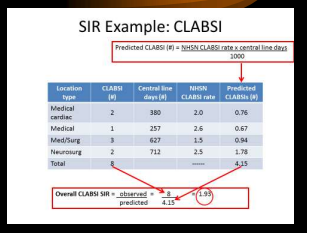
1. Numerator (top number)
   =number of observed infections
2. Denominator (bottom number)
   =number of expected or predicted infections

- Number of predicted infections = calculated based on your hospital's number of procedures, device days, risk factors, nursing units *compared to a standard infection rate* (e.g., historical data, state data, national data)

28

Predicted Number of Infections

- 2015 as baseline year
- Logistic regression/negative binomial regression
- Limited patient level risk adjustment
  - facility type, bed size, med school affiliation, types of units.



Details:
https://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/nhsn-sir-guide.pdf

29

Standardized Infection Ratio

- SIR = # observed infections / # predicted infections
- SIR >1.0 → *more infections than predicted*
- SIR <1.0 → *fewer infections than predicted*
- *~LOWER SIRs are BETTER~*

30

SIR Interpretations

- **SIR=1**
- The number of infections is the same as the number of expected infections
- No progress has been made in reducing infections since the baseline period or compared to another standard population (e.g., all NC, all US).

31

SIR Interpretations

- If the **SIR is less than 1**
  - Fewer infections than predicted based on standard or baseline data
  - Infection reduction/prevention compared to standard or baseline data
  - 1 minus the SIR = percent reduction:
    For example, a SIR of 0.80 means that there was a 20 percent reduction from the standard population or baseline time period

32

SIR Interpretations

- If the **SIR is greater than 1**
  - More infections than predicted based on standard or baseline data
  - Infections are increased compared to standard or baseline data
  - SIR minus 1 = percent increase:
    For example, a SIR of 1.25 means that there was a 25 percent increase from the standard population or baseline time period

33

Example 1: SIR

- CLABSI rate = 4 CLABSI/284 line days
- Predicted Infections = 0.57

- What is the SIR?
- How would you explain the SIR to your administrator?

34

Example 1 SIR: Answers

- CLABSI rate = 4 CLABSI/284 line days
- Predicted Infections = 0.57

- What is the SIR?
  - **SIR = 4 CLABSIs observed/0.57 CLABSIs predicted**
  - **SIR=7.02**
- How would you explain this SIR to your administrator?
  - **We observed more (7 times) CLABSIs than predicted based on comparison to a standard rate***

    ***state the source of standard rate, NHSN? which years?**

35

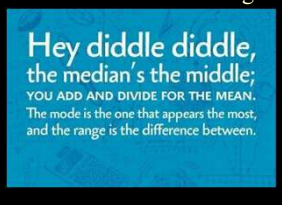
Descriptive Statistics

- Measures of Rates
  - *Rate: How fast disease occurs in a population.*
  - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
  - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion (Variability)
  - *Dispersion: How widely your data is spread from the average.*

36

Measures of Central Tendency

- **Mean:** average of a group of numbers
- **Median:** middle number in an ordered group of numbers
- **Mode:** most common value in a group of numbers



37

Descriptive Statistics

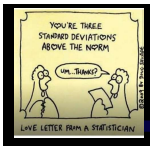
- Measures of Rates
  - *Rate: How fast disease occurs in a population.*
  - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
  - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion (Variability)
  - *Dispersion: How widely your data is spread from the average.*

38

Measures of Dispersion

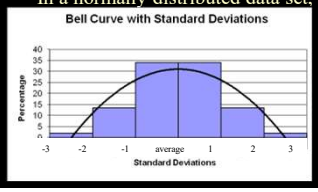
- Range: the largest value minus the smallest value
- Standard deviation: describes the variability or dispersion in the data set

39



Standard Deviation

- **A** measure of degree of variability (spread) in individuals in the sample
  - Standard ("average") deviation ("difference") between an individual's mean and the sample mean
- In a normally distributed data set,



68% of values ± 1 SD

95% of values ± 2 SD

99% of values ± 3 SD

40

Example 2:

- Your administrator is becoming concerned about the impact of healthcare associated infections on the length of stay in your ICU.
- She has asked you to provide her with some data to confirm her suspicions.

41

Example 2:

- Over the last 3 months you have identified a series of 31 ventilator-associated pneumonias with the total length of stay for each ICU patient as follows:

9, 7, 14, 11, 12, 22, 15, 10, 29, 16, 11, 7, 5, 12, 17, 25, 14, 14, 15, 23, 20, 11, 12, 18, 19, 11, 8, 6, 84, 12, 11

42

Example 2:

- What is the:
  - Mean?
  - Median?
  - Mode?
  - Range?

HINT: 5, 6, 7, 7, 8, 9, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12, 14, 14, 14, 15, 15, 16, 17, 18, 19, 20, 22, 23, 25, 29, 84

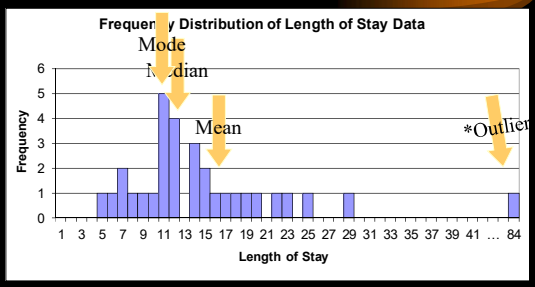
43

Example 2: Answers

- What is the:
  - Mean?
    - 16.1
  - Median?
    - 12
  - Mode?
    - 11
  - Range?
    - 79 *(84[max]-5[min])*
  - Standard Deviation?
    can use programs like Excel to calculate
    - 13.8

44

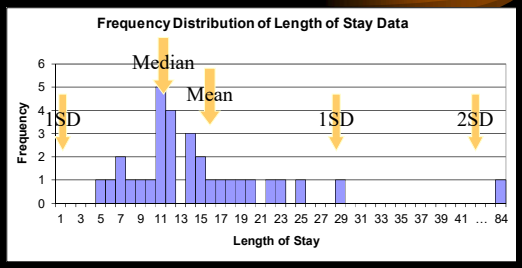
Example 2: Central Tendency



*outlier: a value that falls outside the overall pattern.

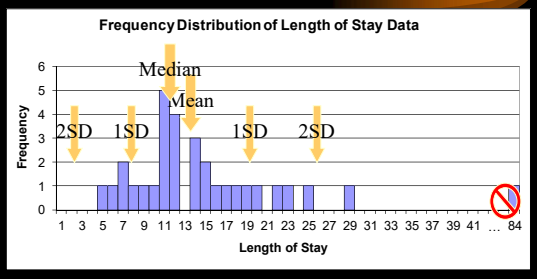
45

Example 2: Dispersion



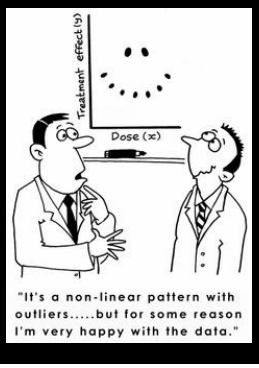
46

Example 2: Dispersion



47

Displaying Surveillance Data



48

Displaying and Interpreting Surveillance Data

- Graphs: a visual representation of data on a coordinate system (e.g., two axes)
- Tables: a set of data arranged in rows and columns

49

Data Types

- Quantitative variables: numerical values
  - *(e.g., number of infections, number of patients)*
- Categorical variables: descriptive groups or categories
  - *(e.g., units in the hospitals, occupational groups)*

50

Features of Graphs and Tables

Graphs and tables should be self-explanatory!

- Clear, concise title: describes person, place, time
- Informative labels: axes, rows, columns
- Appropriate intervals for axes
- Coded and labeled legends or keys
- Use footnotes to:
  - Explain codes, abbreviations, and symbols
  - Note exclusions
  - Note data source

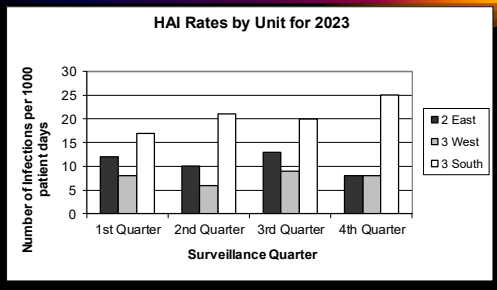
51

Graph Types

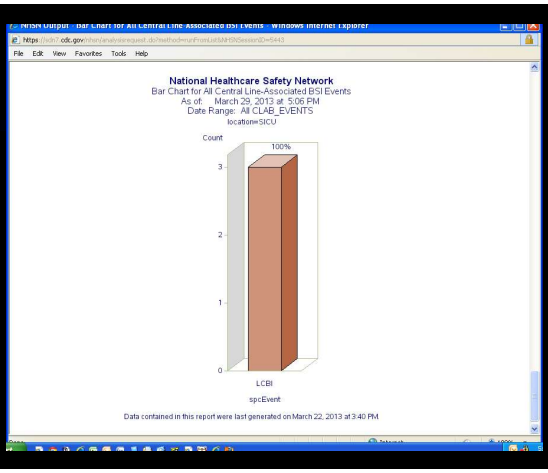
- Bar Graphs
  - *E.g, Histograms (shown in previous example)*
  - *E.g., Comparison between categories*
  - *E.g., Epidemic Curves*
- Line Graphs
  - *E.g., To show trends over time*
- Pie Charts
  - *E.g., As a percentage of a whole*

52

Bar Graph

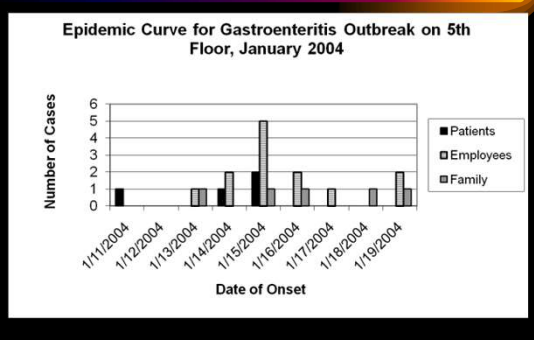


53



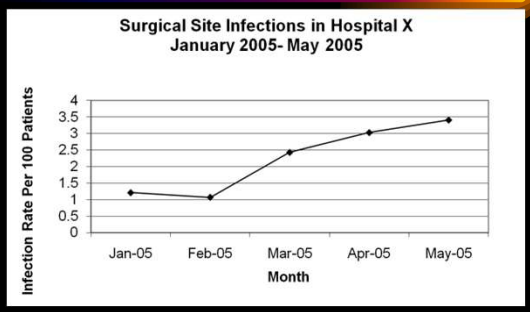
54

Epi Curve



55

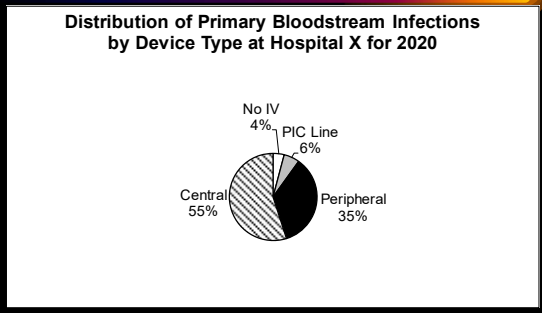
Line Graph



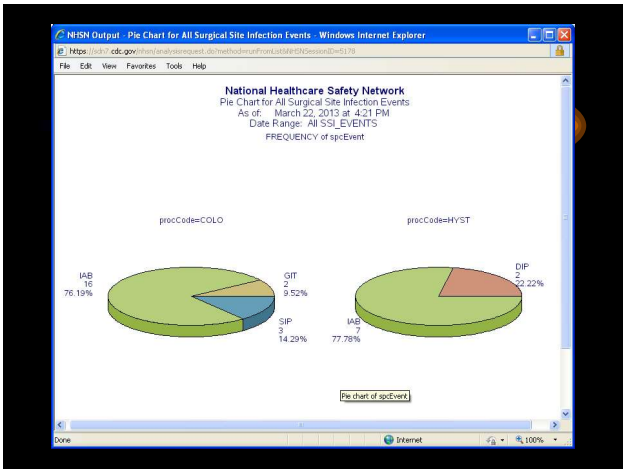
56

57

Pie Chart



58

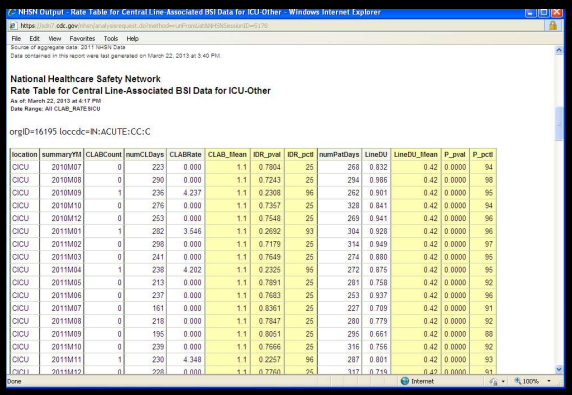


59

Tables

Number of Newly Diagnosed Cases by Age, United States, 2021	
Age Group (Years)	**Number of Cases**
0-4	1242
5-14	1081
15-24	2482
25-44	8153
45-64	10916
65+	7124
Total	**30998**

60


NHSN Output - Rate Table for Central Line-Associated BSI Data for ICU-Other - Windows Internet Explorer

National Healthcare Safety Network
Rate Table for Central Line-Associated BSI Data for ICU-Other

orgID=16195 loccdc=IN:ACUTE:CC:C

location	summaryYM	CLABCount	numCLDays	CLABRate	CLAB_Mean	IDR_pval	IDR_pctl	numPatDays	LineDU	LineDU_Mean	P_pval	P_pctl
CICU	2010M07	0	223	0.000	1.1	0.7804	25	268	0.832	0.42	0.0000	94
CICU	2010M08	0	290	0.000	1.1	0.7243	25	294	0.986	0.42	0.0000	98
CICU	2010M09	1	236	4.237	1.1	0.2308	96	262	0.901	0.42	0.0000	95
CICU	2010M10	0	276	0.000	1.1	0.7357	25	328	0.841	0.42	0.0000	94
CICU	2010M12	0	253	0.000	1.1	0.7548	25	269	0.941	0.42	0.0000	96
CICU	2011M01	1	282	3.546	1.1	0.2692	93	304	0.928	0.42	0.0000	96
CICU	2011M02	0	298	0.000	1.1	0.7179	25	314	0.949	0.42	0.0000	97
CICU	2011M03	0	241	0.000	1.1	0.7649	25	274	0.880	0.42	0.0000	95
CICU	2011M04	1	238	4.202	1.1	0.2325	95	272	0.875	0.42	0.0000	95
CICU	2011M05	0	213	0.000	1.1	0.7891	25	281	0.758	0.42	0.0000	92
CICU	2011M06	0	237	0.000	1.1	0.7683	25	253	0.937	0.42	0.0000	96
CICU	2011M07	0	161	0.000	1.1	0.8361	25	227	0.709	0.42	0.0000	91
CICU	2011M08	0	218	0.000	1.1	0.7847	25	280	0.779	0.42	0.0000	92
CICU	2011M09	0	195	0.000	1.1	0.8051	25	295	0.661	0.42	0.0000	88
CICU	2011M10	0	239	0.000	1.1	0.7666	25	316	0.756	0.42	0.0000	92
CICU	2011M11	1	230	4.348	1.1	0.2257	96	287	0.801	0.42	0.0000	93
CICU	2011M12	0	228	0.000	1.1	0.7760	25	317	0.719	0.42	0.0000	91

61

Interpreting Surveillance Data

62

NHSN data summary, 2013

Urinary catheter associated UTI Rate*					Percentile				
Types of Location: Critical care units	No. Locations	No. of CAUTI	Urinary catheter days	Pooled Mean	10%	25%	50% (median)	75%	90%
Medical cardiac	384	1494	658,345	2.3	0.0	0.7	1.9	3.4	4.9
Medical/surgical ≤15 beds	1645	2429	1,910,118	1.3	0.0	0.0	0.4	1.7	3.1
Surgical cardiothoracic	453	1715	942,852	1.8	0.0	0.7	1.5	2.4	3.4

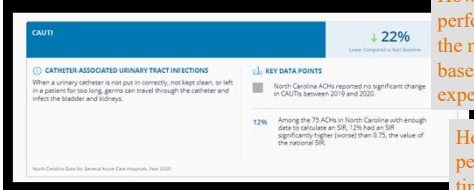
* Number of CA UTIs / Number of urinary catheter days x 1000

63

What does this NHSN data summary tell you?

- What is the mean UTI rate in the medical cardiac critical care unit?
  - 2.3 UTIs per 1000 urinary catheter days
- If your medical/surgical (≤15 beds) critical care unit has a rate of 1.6 UTIs per 1000 urinary catheter days—between what percentiles is it compared to the NHSN data?
  - Between the 50th-75th percentiles
- If your surgical cardiothoracic critical care unit has a rate of 4.2 UTIs per 1000 urinary catheter days—between what percentiles is it compared to the NHSN data?
  - Greater than the 90th percentile

64



How does NC perform compared the national baseline experience?

How does NC perform over time?

ID Year	Year	ID HAI	HAI	State	Standardized Infection Ratio	Predicted Infections	Observed Infections	Number of Facilities Reporting	# Facilities with 1 or more Predicted Infections	% Hospitals Reporting SIR > National SIR
2020	2020	CAUTI	CAUTI	North Carolina	0.78	845.12	659	101	75	12
2019	2019	CAUTI	CAUTI	North Carolina	0.75	840.57	628	98	98	7
2018	2018	CAUTI	CAUTI	North Carolina	0.92	849.26	785	96	71	11
2017	2017	CAUTI	CAUTI	North Carolina	0.89	900.5	805	98	74	11
2016	2016	CAUTI	CAUTI	North Carolina	0.91	907.2	824	94	74	8

https://www.cdc.gov/hai/data/portal/progress-report.html#2018
By state: https://arpsp.cdc.gov/profile/geography

65

Determine the Significance of Changes to Surveillance Data

66

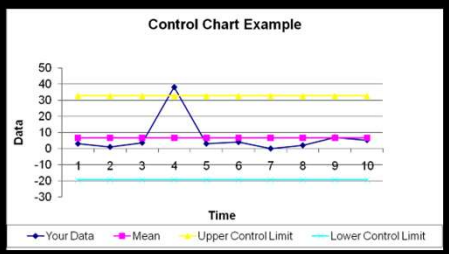
Determine the Significance- How?

- Practical Significance vs. Statistical Significance
- Make comparisons
  - *For example: over time, to other units, to other hospitals (NHSN data)*
  - Remember to choose appropriate data for comparison (*i.e., same denominator units)*
- Apply a type of statistical test
  - *e.g., control charts (for time trends)*
- Other statistical tests and measures
  - P-values
  - 95% confidence intervals

67

Control Charts

- Tool to determine when infection rates are out of range. *How high is TOO high?*



68

Control Chart Example 3:

Month	2015 BSI Rate	Moving Range
1	4.5	--
2	3.2	1.3
3	3.6	0.4
4	3.5	
5	3.0	
6	4.0	
7	4.1	
8	4.6	
9	4.8	
10	5.2	
11	5.7	
12	6.5	

- Find the mean of the BSI rates for the last year.
- Calculate the moving ranges (subtract month 1 from 2, month 2 from 3…) and take absolute values (no negative values).
- Calculate the mean of the moving ranges.

69

Control Chart Example 3: Answers

Month	2015 BSI Rate	Moving Range
1	4.5	
2	3.2	1.3
3	3.6	0.4
4	3.5	0.1
5	3.0	0.5
6	4.0	1.0
7	4.1	0.1
8	4.6	0.5
9	4.8	0.2
10	5.2	0.4
11	5.7	0.5
12	6.5	0.8

- Find the mean of the BSI rates.
  =4.4
- Calculate the moving ranges
  See table
- Calculate the mean of the moving ranges.
  =0.5

70

Control Chart Example 3:

- Calculate Upper Control limit= Mean + (2.66 x Mean of Moving Range)
- Calculate Lower Control limit= Mean - (2.66 x Mean of Moving Range)
- Draw horizontal lines at the mean, UCL and LCL based on your historical data
- Then graph your current data and use the limits to identify potential problems.

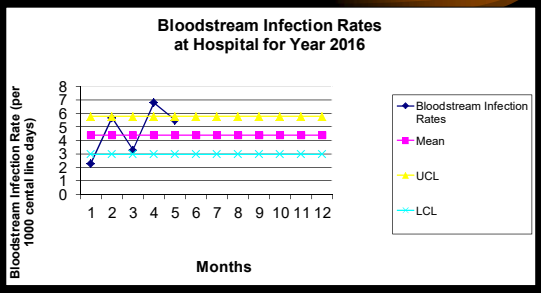
71

Control Chart Example 3: Answers

- Calculate Upper Control limit= 5.8
- Calculate Lower Control limit= 3.0
- Draw horizontal lines at the mean, UCL and LCL based on your historical data
- Then graph your current data and use the limits to identify potential problems.

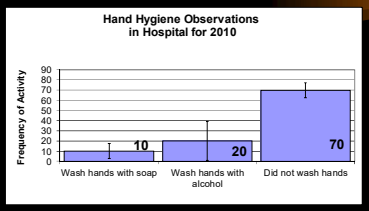
72

Control Chart Example 3:



73

Samples of P values and 95% Confidence Intervals in use



"Our study showed that people who washed their hands were less likely to get sick (P=0.06) and more likely to be nurses (P=0.01)."

74

Statistical Tests – Why do we need them?

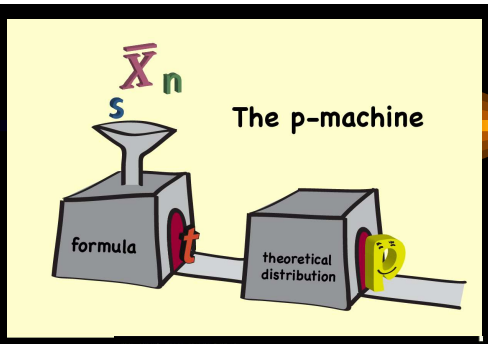
- **Is this real?** Use in clinical research designed to tell if the difference seen is due to chance, or due to some other cause (i.e. a real difference)
- We use these measures to make an **inference**
  - Process of drawing a conclusion about a larger group based on a sample or subset of the group

75

P value

- **P value**: probability of finding a difference as extreme or more extreme than what was found, assuming that the null hypothesis is true
  - Can be used as a measure of the degree of compatibility between observed data and null hypothesis
  - The conventional (yet arbitrary) threshold is 0.05, below which the null hypothesis is rejected
  - 0.05 accepts a 5% risk of a Type 1 error

76



$\bar{x}$ average value
s standard deviation
n sample size (number of measurements)
t test statistic = $\frac{\bar{x}}{s/\sqrt{n}}$
p p-value (probability)

77

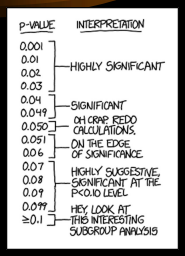
P-Value Example:

- "Our study showed that people who washed their hands were less likely to get sick (P=0.06) and more likely to be nurses (P=0.01)."

78

P-Value Interpretation

- Probability that the difference does not reflect a true difference and is only due to chance.
- e.g., P=0.05 means that 95 out of 100 times your estimate was truly significant
- Generally a level of P<0.05 is considered "statistically significant."



79

Estimation: 95% Confidence Interval

- **95% Confidence Interval (CI)**: calculated range of values surrounding the point estimate that are consistent with true effect
  - Formula: point estimate of the mean +/- $(2^* {}^{s}/_{\sqrt{n}})$
- Means that you are 95% confident that the true average value lies within this interval.

80

Statistical Variation of Estimates

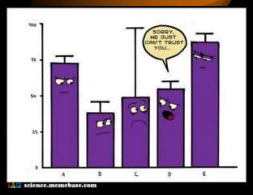
- Consider your calculated infection rate to be an estimation of the true rate.

  Why an estimation?
  - You may only do surveillance on a sample of patients in your hospital.
  - If surveillance activities were repeated by other IPs, your numerators may vary slightly based on interpretation of case definitions, available clinical information in the chart, etc.

81

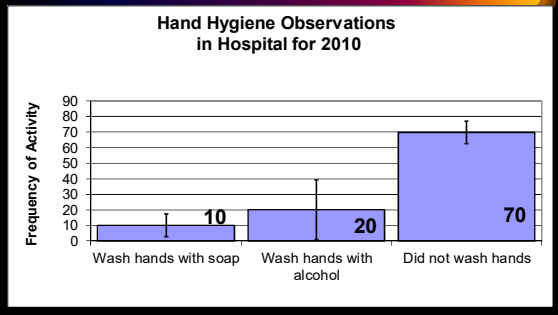
95% Confidence Interval Interpretation

- Confidence interval size:
  - Wide: less confident with that estimate
  - Narrow: more confident with that estimate
- For comparisons,
  - Overlapping intervals suggest no significant difference
  - Non-overlapping intervals suggest significant differences



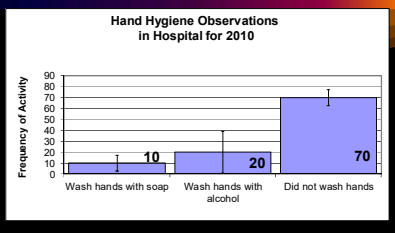
82

95% Confidence Interval Example:



83

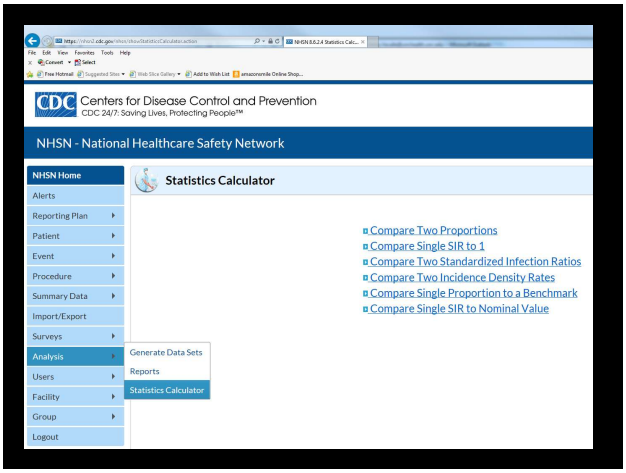
95% Confidence Interval Example:



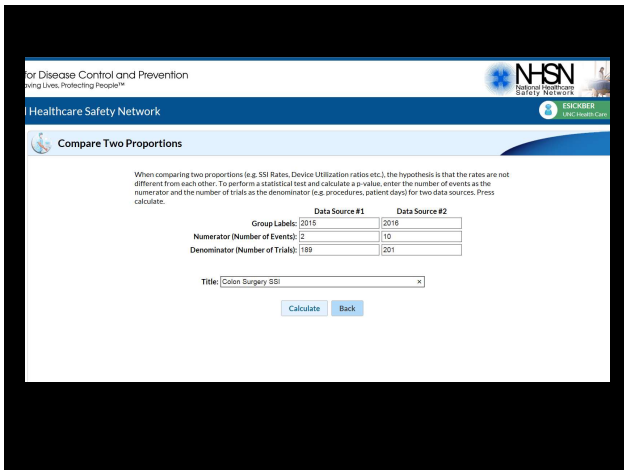
Is the frequency of not washing hands at this hospital statistically significantly different than the frequency of washing hands with soap? *YES – the 95% CI do not overlap*

Is the frequency of washing hands with soap at this hospital statistically signficantly different than the frequency of washing hands with alcohol? *NO– the 95% CI overlap*

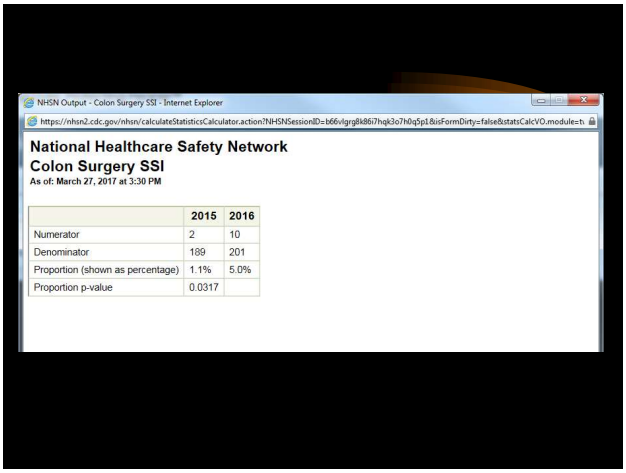
84



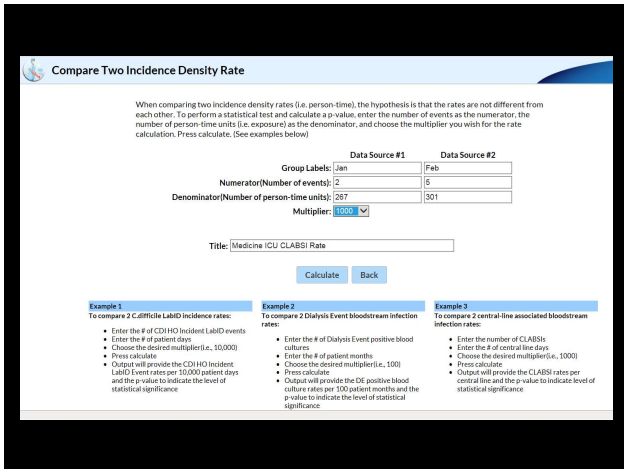
85



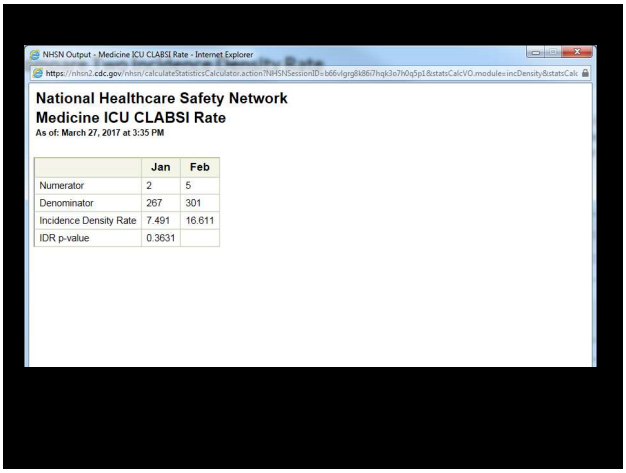
86



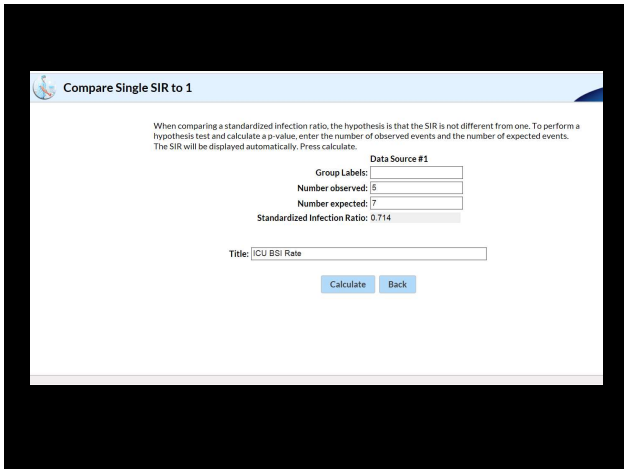
87



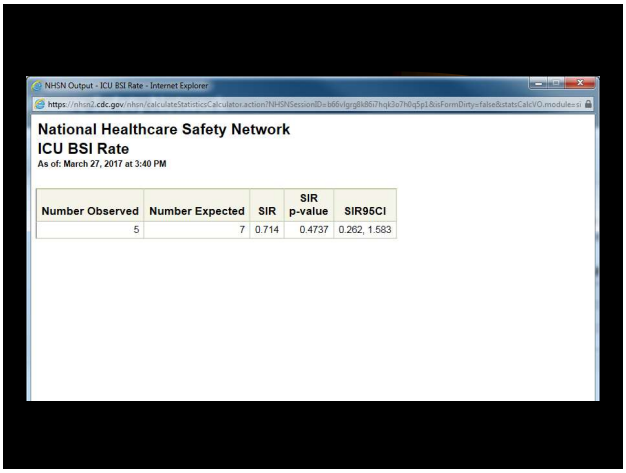
88



89



90



91

Conclusions

- Describe Surveillance Data
- Display and Interpret Surveillance Data
- Determine the Significance of Changes to Surveillance Data

92

Questions?

93

Group Exercises Using Excel

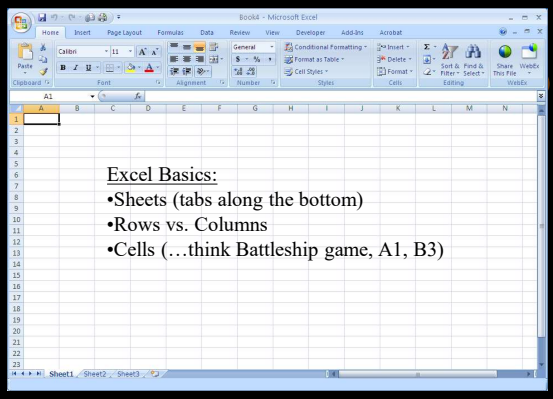
- Infection Rates
  - Create a table
  - Practice formulas
  - Optional activities
    - Graph rates
    - Add 2nd series on graph for NHSN benchmark
    - SIR calculation

94

Group Exercises Using Excel

- Outbreak Investigation
  - Create line-listing of outbreak cases
  - Practice formatting cells, copy/paste, sorting
  - Optional activities:
    - Create a frequency table of cases
    - Graph outbreak epi-curve

95



96

Exercise Wrap-up

- Use Excel as a tool for
  - Calculations of infection rates
  - Creating line-listing for outbreaks or cluster investigations
  - Displaying data graphically
- Use each cell in Excel to capture single piece of data
- Graphs and tables should be self-explanatory!
  - Clear, concise title, informative labels
- Practice, practice, practice!

97