Interpreting Uncertainty: Making Sense of Statistics

Daniel Bronson-Lowe, PhD, CIC
Carle Hospital and Physician Group
Hypotheses

- **Null hypothesis**: values are equal.

- **Alternative hypothesis**: values differ.

- These statements are mutually exclusive.
  - They cover all possible outcomes.
  - In the end, only one can be selected.
Hospital A vs. Hospital B

- **Null hypothesis:**
  - The average levels of hand hygiene compliance at Hospital A and Hospital B are the same.

- **Alternative hypothesis:**
  - The average levels of hand hygiene compliance at Hospital A and Hospital B are different.
Hand Hygiene Compliance

Hospital A: 54%

Hospital B: 86%
Hand Hygiene Compliance

Hospital A

54%

Hospital B

86%
Hand Hygiene Compliance

Hospital A

54%

69%

Hand Hygiene Compliance

Hospital B

69%

86%

Hand Hygiene Compliance
## Error Options

<table>
<thead>
<tr>
<th>Reality</th>
<th>Same (null)</th>
<th>Different (alt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same (null)</td>
<td>Correct</td>
<td>False Positive</td>
</tr>
<tr>
<td>Different (alt)</td>
<td>False Negative</td>
<td>Correct</td>
</tr>
</tbody>
</table>

### The False **Positive** Scenario (Type I Error)

**Your Conclusion**
- There is a difference.
  - (Alternative Hypothesis)

**Reality**
- There is NO difference.
  - (Null Hypothesis)
The probability of concluding a difference is real when it is actually just random variation (a false positive).
$\alpha = \text{Probability of a False Positive}$

<table>
<thead>
<tr>
<th>Chance of False Positive</th>
<th>1 in 100</th>
<th>1 in 20</th>
<th>1 in 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

$\alpha = 0.01$  
$\alpha = 0.05$  
$\alpha = 0.10$
**p-value**: the probability that this difference (or a more extreme one) was caused by random chance if the null hypothesis is true.
Risk of a False Positive

<table>
<thead>
<tr>
<th>Alpha</th>
<th>vs.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The amount of risk you are willing to accept.</td>
<td>The amount of risk present.</td>
<td></td>
</tr>
</tbody>
</table>
\( p = 0.08 \)

- \( p \)-value > \( \alpha \)
- The risk of a false positive is too high.
- Conclude there is no difference.
Significant

Chance of False Positive

0%  5%  10%

\( p = 0.02 \)

- p-value < \( \alpha \)
- The risk of a false positive is acceptable.
- Conclude a difference exists!
Standardized Infection Ratios

$$SIR = \frac{\text{Observed infections}}{\text{Expected infections}}$$

Better

Same

Worse

0 1 2
SIR
Standardized Infection Ratios

- **Null hypothesis:**
  - Your infection rate and the benchmark infection rate are the same.

- **Alternative hypothesis:**
  - Your infection rate and the benchmark infection rate are different.
## Standardized Infection Ratios

### CABG-Related Surgical Site Infections

<table>
<thead>
<tr>
<th>Facility</th>
<th>Number of Procedures Performed</th>
<th>Number of Infections</th>
<th>Expected Number of Infections</th>
<th>SIR</th>
<th>SIR p-value</th>
<th>95% Confidence Interval for SIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>290</td>
<td>1</td>
<td>5.8</td>
<td>0.2</td>
<td>0.03</td>
<td>0.0, 0.9</td>
</tr>
<tr>
<td>B</td>
<td>80</td>
<td>4</td>
<td>1.5</td>
<td>2.6</td>
<td>0.65</td>
<td>0.7, 6.7</td>
</tr>
<tr>
<td>C</td>
<td>1500</td>
<td>75</td>
<td>28.5</td>
<td>2.6</td>
<td>0.01</td>
<td>2.1, 3.3</td>
</tr>
</tbody>
</table>

\( \alpha = 0.05 \)
Standardized Infection Ratios

$\alpha = 0.05$

<table>
<thead>
<tr>
<th>SIR</th>
<th>SIR p-value</th>
<th>95% Confidence Interval for SIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.03</td>
<td>0.0, 0.9</td>
</tr>
<tr>
<td>2.6</td>
<td>0.65</td>
<td>0.7, 6.7</td>
</tr>
<tr>
<td>2.6</td>
<td>0.01</td>
<td>2.1, 3.3</td>
</tr>
</tbody>
</table>
Sample Size

Small

Large

All
Data Variability

Large Variability

Small Variability
Large Variability

Hand Hygiene Compliance

Hospital A

54%

Hand Hygiene Compliance

Hospital B

86%

Hand Hygiene Compliance
Little Variability

Hand Hygiene Compliance

Hospital A

54%

Hand Hygiene Compliance

Hospital B

86%
<table>
<thead>
<tr>
<th>Alpha ($\alpha$)</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>99% CI</td>
</tr>
<tr>
<td>0.05</td>
<td>95% CI</td>
</tr>
<tr>
<td>0.10</td>
<td>90% CI</td>
</tr>
</tbody>
</table>
Error Options

<table>
<thead>
<tr>
<th>Reality</th>
<th>Same (null)</th>
<th>Different (alt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same (null)</td>
<td>Correct</td>
<td>False Positive</td>
</tr>
<tr>
<td>Different (alt)</td>
<td>False Negative</td>
<td>Correct</td>
</tr>
</tbody>
</table>

The False **Negative** Scenario (Type II Error)

**Your Conclusion**

There is NO difference.

(Null Hypothesis)

**Reality**

There is a difference.

(Alternative Hypothesis)
Changing Alpha...

More Likely

False Positive

False Negative

Less Likely
What Drives Power?

- Difference of interest
- Sample size
- Data variability
- Level of significance ($\alpha$)
- Test used
Clinical Relevance

- Statistical significance ≠ Clinical relevance

Facility A
p = 0.34

Facility B
p = 0.01

SIR
<table>
<thead>
<tr>
<th>Vaccine, # Doses</th>
<th>Relative Risk</th>
<th>95% Confidence Interval for RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTaP, 3</td>
<td>1.34</td>
<td>1.32 - 1.35</td>
</tr>
<tr>
<td>IPV, any</td>
<td>1.01</td>
<td>1.01 - 1.01</td>
</tr>
<tr>
<td>IPV, 3</td>
<td>1.27</td>
<td>1.25 - 1.29</td>
</tr>
<tr>
<td>PCV, any</td>
<td>1.04</td>
<td>1.03 - 1.04</td>
</tr>
<tr>
<td>PCV, 3</td>
<td>1.37</td>
<td>1.35 - 1.39</td>
</tr>
</tbody>
</table>

Garbage In, Garbage Out

Bad Data

Wrong Statistical Test
- t-test
- ANOVA
- Chi-square
- etc.

Test Statistic
(e.g. t, F, X^2)

More Math

p-value
Nothing is Certain

- Statistically significant result?
  - Alpha = 0.05
  - Wrong 1 time in 20

- No significant finding?
  - Power = 80%
  - Wrong 1 time in 5
A Couple of Examples...
Example: CHG & NICU CLABSI

- Implemented use of chlorhexidine gluconate (CHG) bathing to reduce CLABSI rates in NICU

- Compared mild soap to CHG:
  - CHG bathing regimens varied by birth weight and chronological age.

- Compared Incidence Rate Ratios (IRR)
  - IRR = 1.0 means the two groups are the same

Example: CHG & NICU CLABSI

- “CLABSI rates decreased during the intervention period for CHG–bathed neonates.”
  - IRR = 0.33
  - 95% Confidence Interval = 0.15–0.73

Is this a statistically significant finding?
A) Yes
B) No
“CLABSI rates decreased during the intervention period for CHG–bathed neonates.”
- IRR = 0.33
- 95% Confidence Interval = 0.15–0.73

What was the value of alpha for this study?
A) 0.01
B) 0.05
C) 0.10
D) 0.50
Example: CHG & NICU CLABSI

“CLABSI rates decreased during the intervention period for CHG–bathed neonates.”
- IRR = 0.33
- 95% Confidence Interval = 0.15–0.73

What was the likelihood of a false positive result?

A) 1%
B) 5%
C) 20%
D) 95%
Example: CHG & NICU CLABSI

“CLABSI rates decreased during the intervention period for CHG–bathed neonates.”
- IRR = 0.33
- 95% Confidence Interval = 0.15–0.73

Which of these is a believable p-value for this comparison?
A) 0.03  
B) 0.06  
C) 0.12  
D) 0.83
## Example: HCW Flu Vaccination

<table>
<thead>
<tr>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaccination Rate</td>
<td>Vaccination Rate</td>
</tr>
<tr>
<td>50% → 95%</td>
<td>50% → 95%</td>
</tr>
</tbody>
</table>

\[ p\text{-value} = 0.120 \quad \alpha = 0.05 \quad \text{p-value} = 0.001 \]

Which study had a statistically significant result?
### Example: HCW Flu Vaccination

<table>
<thead>
<tr>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaccination Rate</td>
<td>Vaccination Rate</td>
</tr>
<tr>
<td>50% → 95%</td>
<td>50% → 95%</td>
</tr>
</tbody>
</table>

- **Study 1**
  - **p-value = 0.120**
  - **α = 0.05**

- **Study 2**
  - **p-value = 0.001**

**What might be making that one significant?**

A) The researchers used a more expensive statistician.
B) The researchers were more willing to risk a false positive.
C) It’s more clinically relevant.
D) It had a larger sample size.
Study Design
Study Designs

Observational Studies

Cross-sectional Surveys

Case-Control Studies

Cohort Studies
Cross-sectional Survey

- Survey a random sample at a single point in time
Cross-sectional Survey

- **Pros**
  - Quick and cheap (relatively)
  - Can cover an entire population

- **Cons**
  - Based primarily on self-report
  - Lack of a time sequence
  - Getting a suitable sample can be difficult
Watch people over a period of time and compare outcomes.
Watch CABG patients for 1 year to see if those with BMI $\geq 30$ are at higher risk for surgical site infection.
Cohort Study

- **Pros**
  - Easier and cheaper than randomized controlled trial
  - Can establish a temporal relationship
  - Can watch for multiple outcomes

- **Cons**
  - Expensive
  - Require a large sample size
  - Inefficient for rare outcomes
  - Can take a long time for outcome to occur
  - No randomization
Compare people who have a certain condition to those who do not.
Case–Control Study

- Determine risk factors for CABG surgical site infections occurring over last 6 months.
Case-Control Study

- **Pros**
  - Good for studying rare outcomes
  - Quicker than cohort: the outcome has already occurred
  - Can look at multiple risk factors
  - Useful as an initial study to establish a possible association

- **Cons**
  - Reliance on historical data / memory (risk of recall bias)
  - Examining single outcome
  - Selection of a control group can be difficult
Study Designs

- Experimental Studies
  - Randomized Controlled Trial
  - Quasi-Experimental Studies
Randomized Controlled Trial

- Participants are randomly assigned to an experimental/intervention group or a control group.

- Often used for testing of new drugs and treatments.
Randomized Controlled Trial

Participants

Random allocation

Intervention Group

Follow-up

Control Group

Follow-up
Randomized Controlled Trial

- Is surgical skin prep A better at preventing SSIs than surgical skin prep B?
Randomized Controlled Trial

Pros
- Randomization addresses many types of bias
- Easier to incorporate blinding

Cons
- Ethical concerns
- Results may not be generalizable
- Expensive
- Volunteer bias
Randomized Controlled Trial

- Only studying one population
  - Group allocation should not be based on two populations

- Only difference between the groups should be the variables being studied
  - Additional differences may be confounders.
An intervention occurs, but an element of RCT is missing: Randomization.

Also may be missing:
- Pre-post test design
- Control groups

Often used to assess the impact of a program or intervention (e.g. process improvement projects)
Quasi-Experimental Study

Intervention Group

Intervention

Intervention Group

Control Group

Control Group
Quasi-Experimental Study

- Is surgical skin prep A better at preventing SSIs than surgical skin prep B?
Quasi–Experimental Study

- **Pros**
  - More practical than RCT
  - Can use “natural experiments”

- **Cons**
  - Lack of randomization can lead to bias
  - Difficulty controlling for confounding variables
  - Regression to the mean
  - Maturation effects
If you want to inspire confidence, give plenty of statistics.

It does not matter that they should be accurate, or even intelligible, as long as there is enough of them.

– Lewis Carroll