

Navigating the Data Science Landscape

Half-day workshop at
International Conference on Applied Statistics, ICAS 2019
Institute of Statistical Research and Training (ISRT)
University of Dhaka, Dhaka, Bangladesh
December 26, 2019

Enayetur Raheem, PhD
enayetur.raheem@brfbd.org

Affiliations and Disclosures

- Opinion expressed here are my own and not of the organizations I am affiliated with.
- No funding was received to offer this workshop.
- This presentation may promote **Biomedical Research Foundation**, as we are actively looking for fresh graduate statisticians to involve in our research projects.

Outline

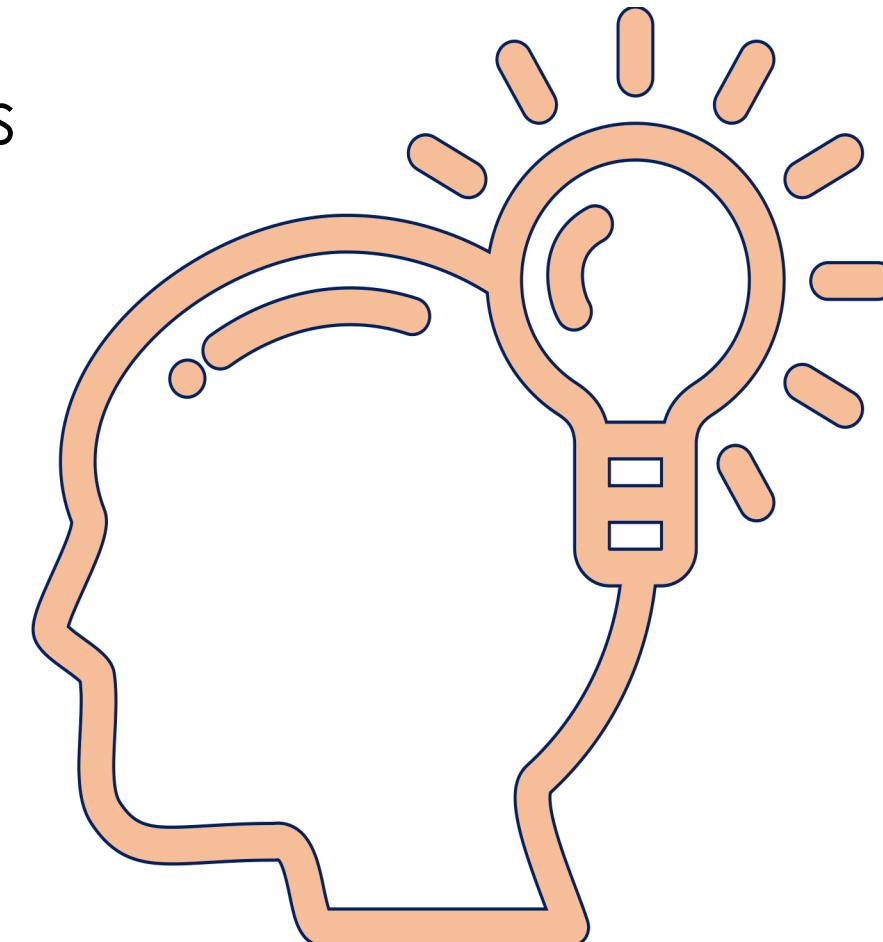
Part 1 Data Science in the Industry	Part 2 How works get done in the Industry	Part 3 Becoming a Data Scientist
<p>Examples from different domains</p> <ul style="list-style-type: none"> • Predictive modeling • Descriptive studies 	<ul style="list-style-type: none"> • How do managers leverage data scientists to solve business problems • A hierarchical view of the org structure and where do data scientists fit 	<ul style="list-style-type: none"> • Resources • Prescription for success • Training/Learning opportunities

Part 1

Use Cases

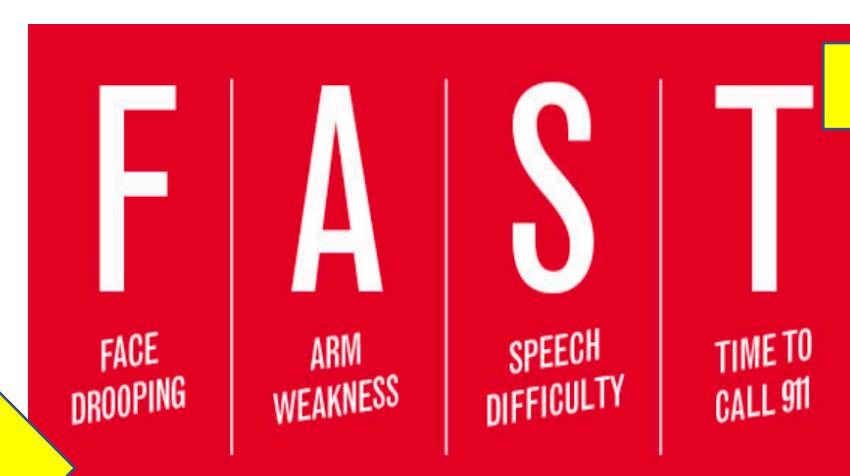
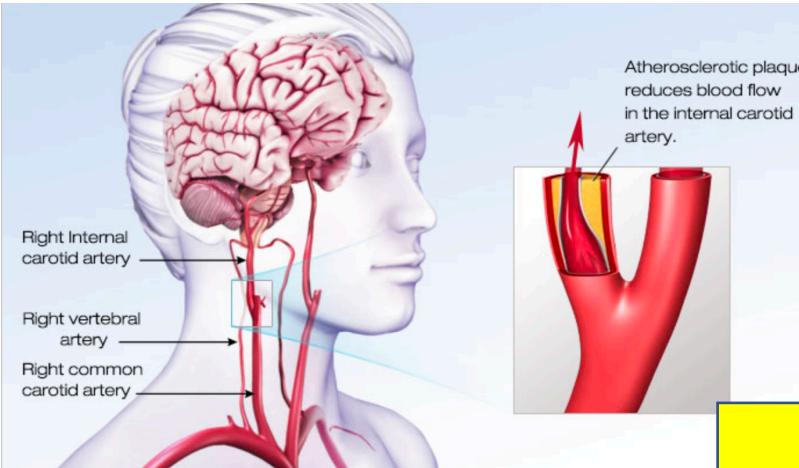
What to keep thinking about

- Someone asked me if statisticians are still relevant in the industry
- Keep thinking about where do you see yourself fit in the process
- Keep thinking about how would **you** attempt to solve it



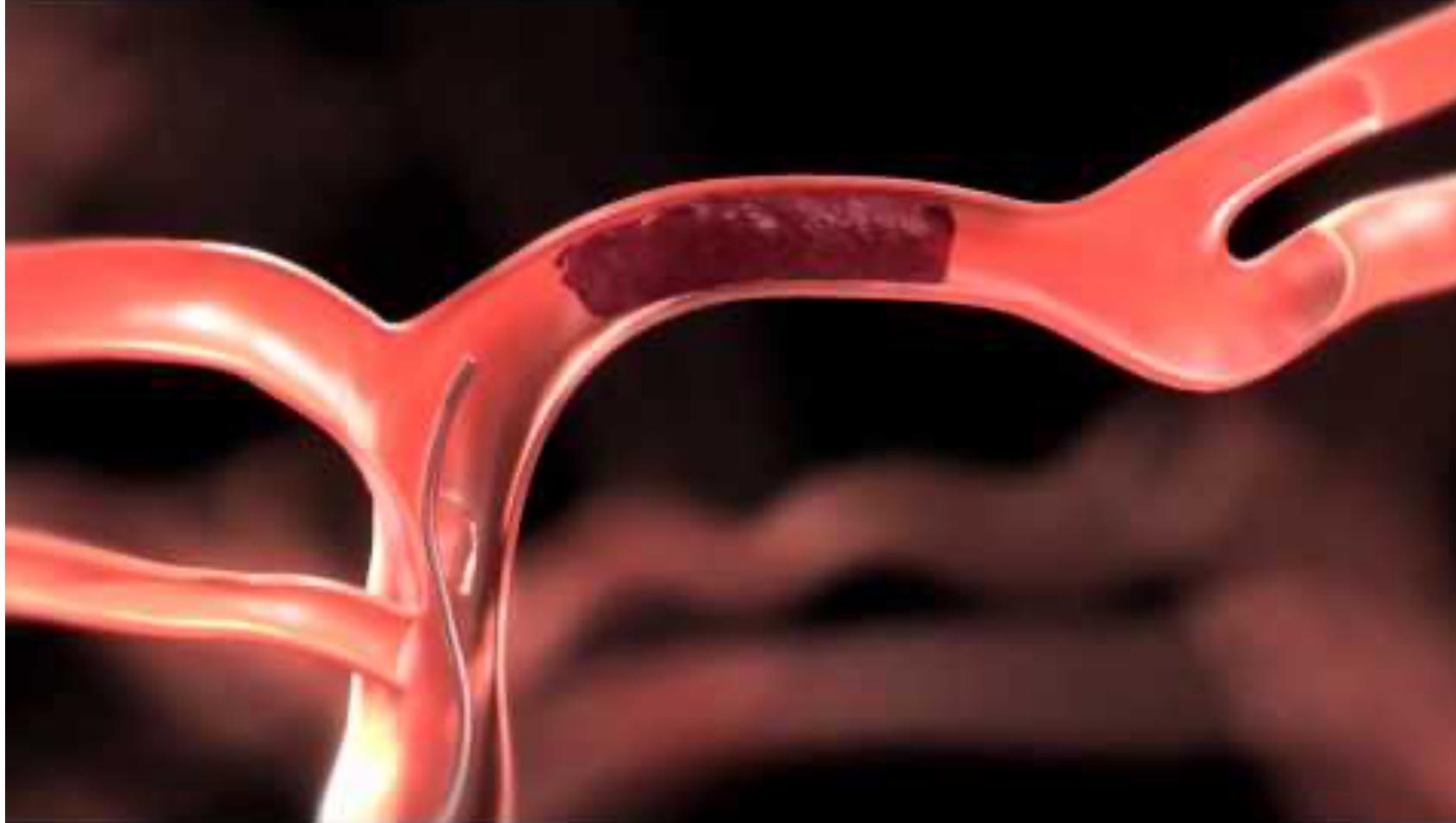
Case 1: Process Improvement

SCENARIO: Patients with suspected stroke arriving at the Emergency Room.



Stroke.org

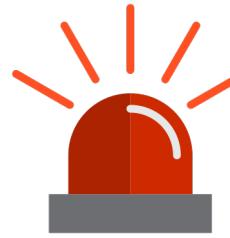
Animation: Mechanical Thrombectomy



https://www.youtube.com/watch?v=Cr9Pb3lSIA&feature=emb_logo

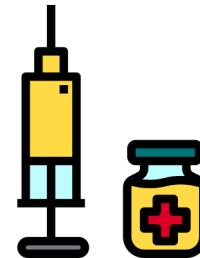


MINUTES



LAST KNOWN WELL TO DOOR
ACTUAL: >60, TARGET: <60

DOOR TO ER PHYSICIAN
ACTUAL: 15, TARGET: < 10



CT REPORT TO IV TPA
ACTUAL: 25, TARGET: <20

CT SCAN TO INTERPRETATION
ACTUAL: >45, TARGET: < 25

ER PHYSICIAN TO CT SCAN
ACTUAL: >35, TARGET: < 25



Case 1: Questions to Consider

- Who will use it and how
- What metrics are most relevant
- How best to present the results
- What are the ultimate goals of doing this? (Business perspective)



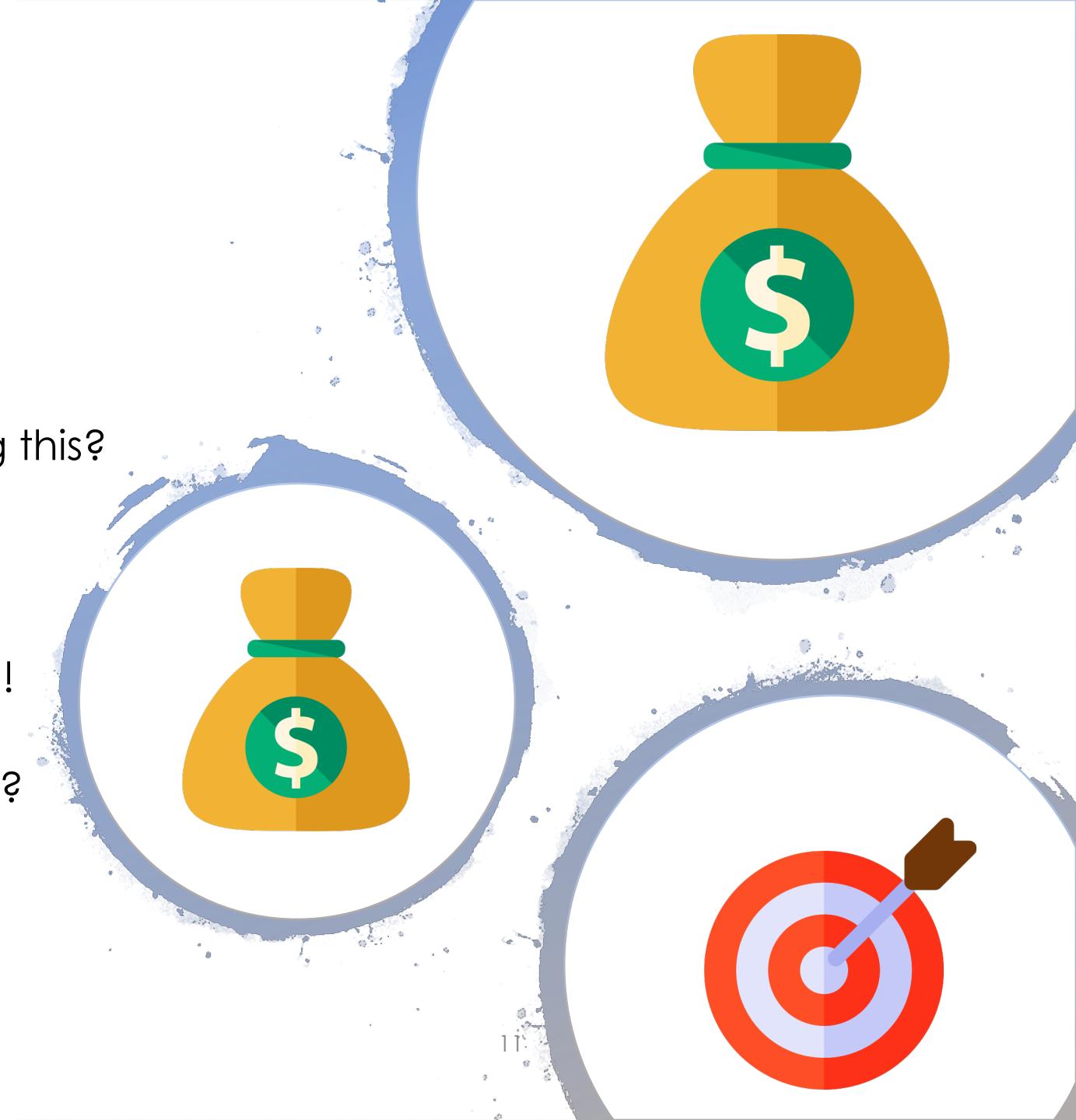
Case 2: Impact Evaluation



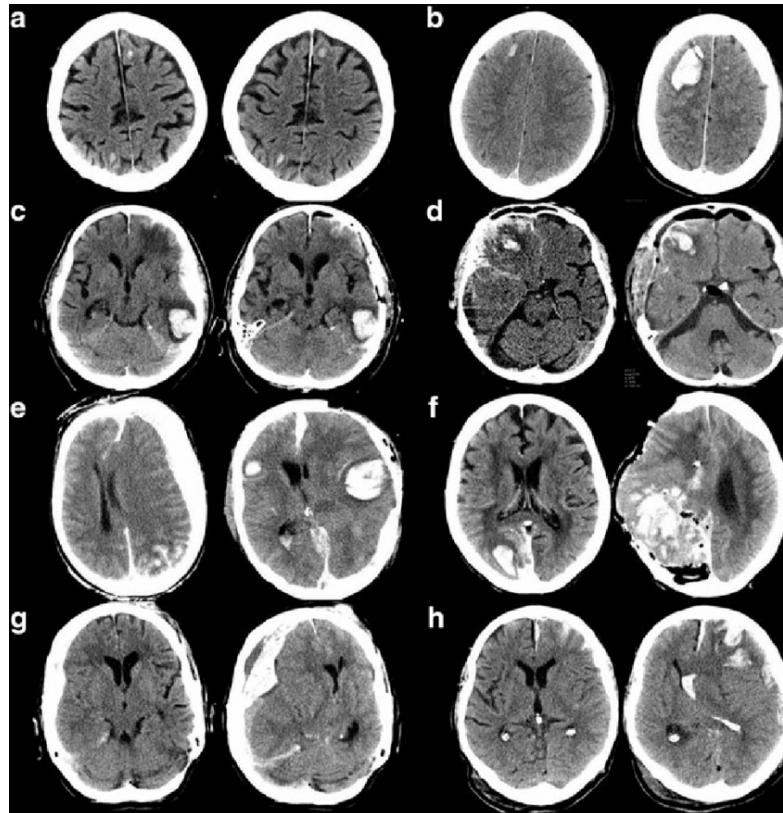
SCENARIO: A county is considering ambulance with CT Scan facility. Costs several million dollars. Should they implement it or not?

Case 2: Questions to Consider

- What are the ultimate goals of doing this? (impact on outcomes, business)
- What metrics are most relevant?
- How to evaluate the intervention?
 - Note, it's a million-dollar question!
 - What study design to use?
 - Where do you get the data from?
 - What data to collect?



Case 3: Traditional Vs (+) AI



SCENARIO: Whether to adopt AI technology over traditional?

https://www.researchgate.net/publication/261747269_Early_Hemorrhagic_Progression_of_Traumatic_Brain_Contusions_Frequency_Correlation_with_Coagulation_Disorders_and_Patient_Outcome_A_Prospective_Study/figures?lo=1
<https://medtechasia.in/ischemaviews-rapid-approved-in-canada/>

Comparison of existing technology vs AI powered technology on some predefined metrics

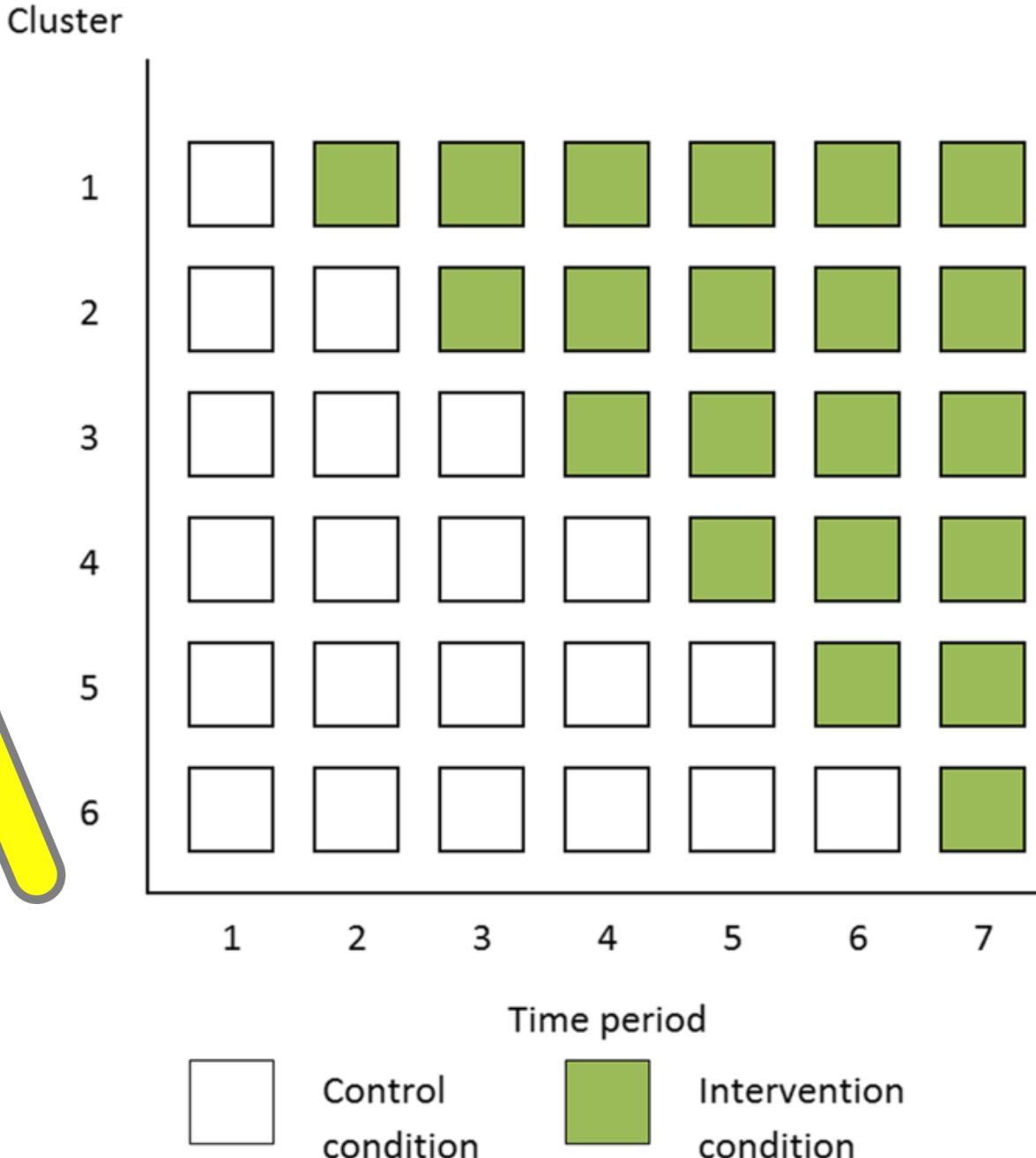
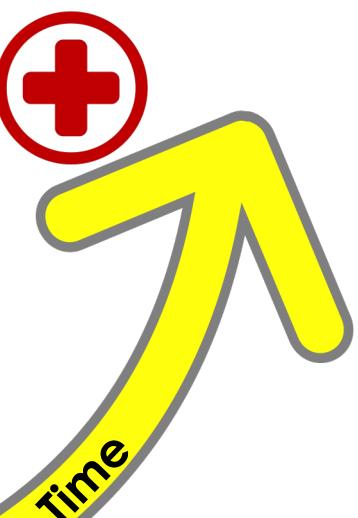
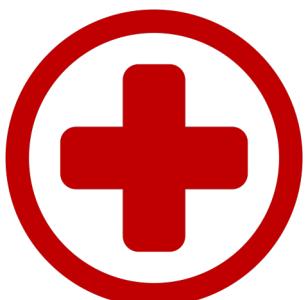
Crossover trial
(subjects are their own control)

Hospitals are the units of measurements

Cluster randomized trial

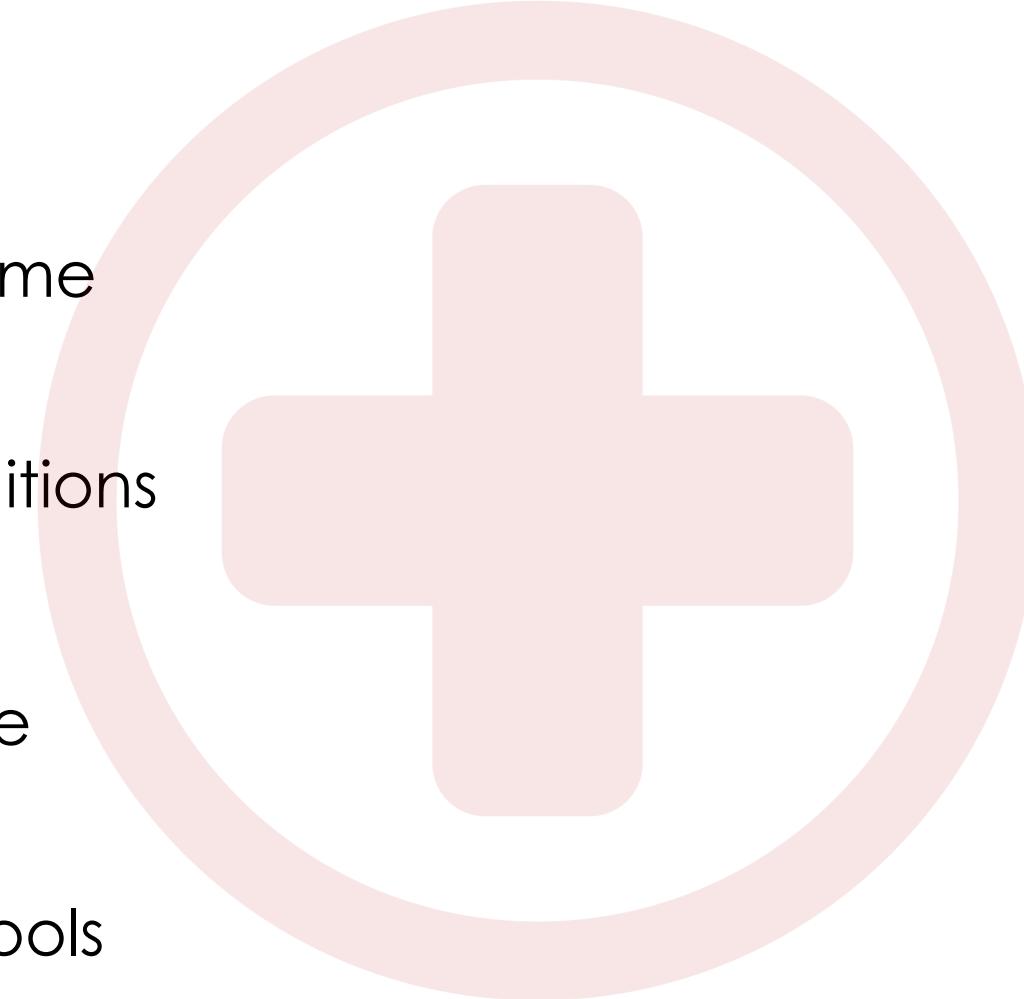
Washout period is hard to ensure because people cannot forget their practice so easily, and unintentional bias will be introduced

Stepped-wedge cluster randomized trial

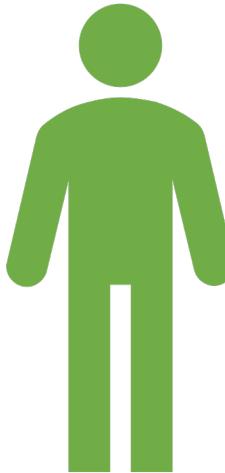


Case 3: Questions to Consider

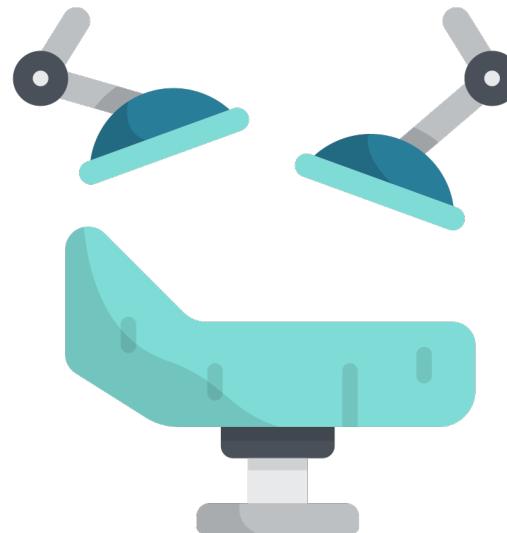
- Choice of a design considering
 - Feasible sample size given the time constraint
 - Budget
 - Implementation in varying conditions with varying groups of people
- Feasibility of the design
- How would the result help make the decision
- Implementation challenges due to potential replacement of existing tools



Case 4: Predicting Complications after Surgery



Patients to undergo elective surgery



Likelihood of complications after surgery



[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Informed decision/preparation before surgery



Enter Patient and Surgical Information



Please enter as much of the following information as you can to receive the best risk/benefit estimates. A rough estimate will still be generated if you cannot provide all of the information below.

Procedure Types: Band Lap Sleeve Lap Bypass BPDDS

BMI Calculation:

Height: 70 in / 178 cm

Weight: lb / kg

Age: 55 Sex: Male Hispanic Ethnicity:

Male No

Race:

Asian

ASA Class:

I. Healthy Patient

Diabetes:

No

Functional Status:

Independent

Current Smoker within 1-year

Sleep Apnea

History of PE

Cardiac Risk

Vascular Risk

History of Severe COPD

Hypertension requiring medication

Hyperlipidemia

GERD

Dialysis

Previous Foregut Surgery

Steroid Use for Chronic Condition

Reset All Selections

Compute Results

<https://riskcalculator.facs.org/bariatric/patientoutcomes.jsp>

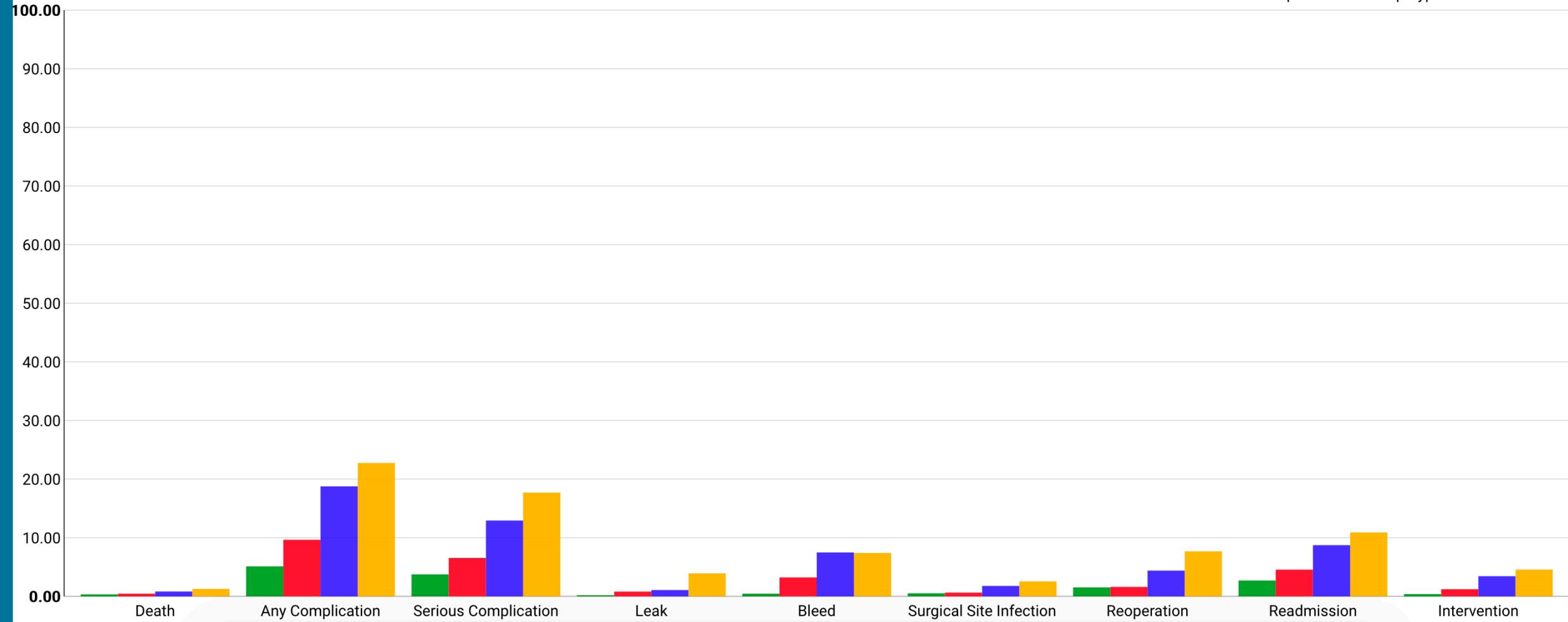


Risk Factors: 43.04 (BMI), 55.00 (Age), Male, Asian (Race), No (Hispanic Ethnicity), Healthy Patient, Smoker, Sleep Apnea, Hyperlipidemia, COPD, Vascular, Cardiac

Outcomes

View Enlarged Chart

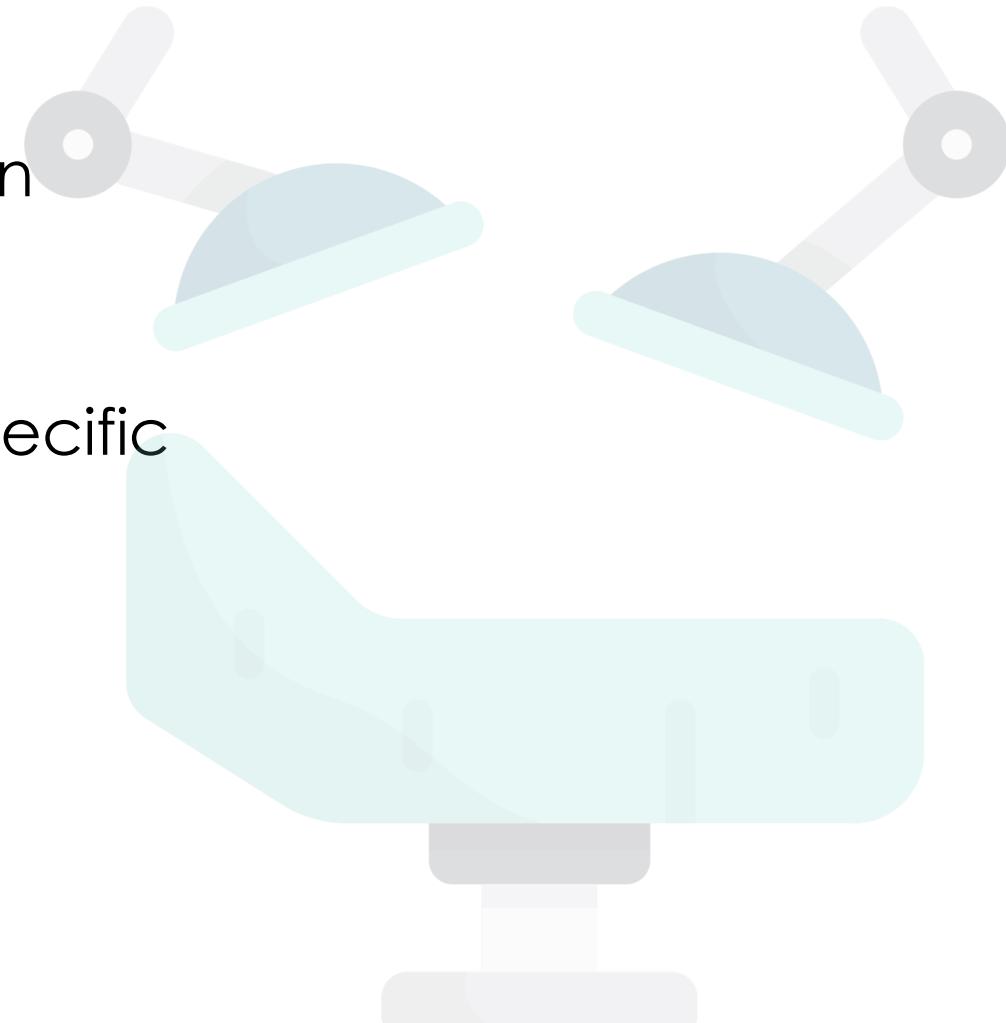
● Band ● Lap Sleeve ● Lap Bypass ● BPDDS



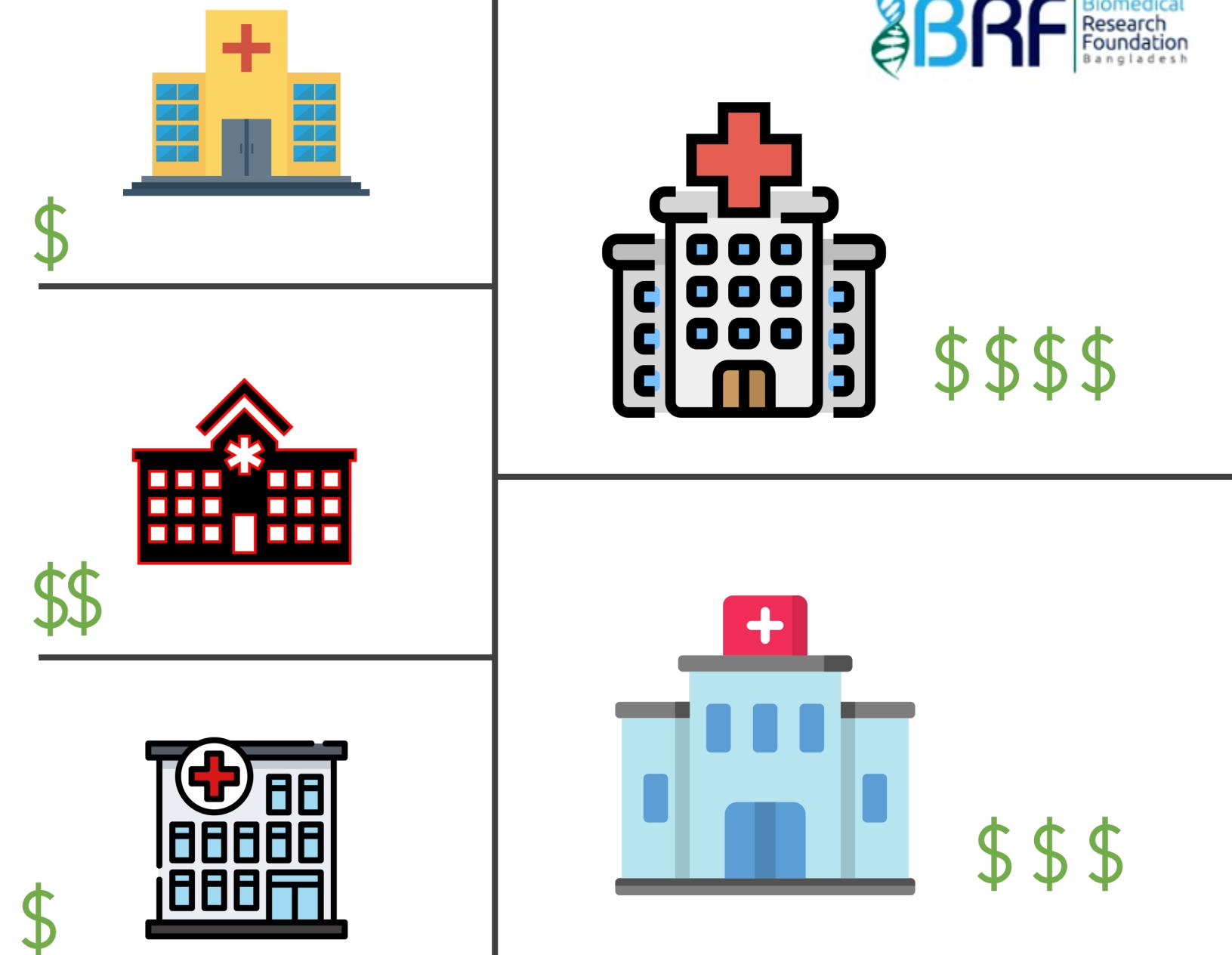
Case 4: Questions to Consider

- Why do we need another prediction model if there's one already there
- Where do we get the data from?
- Can we build specific models for specific type of surgery?

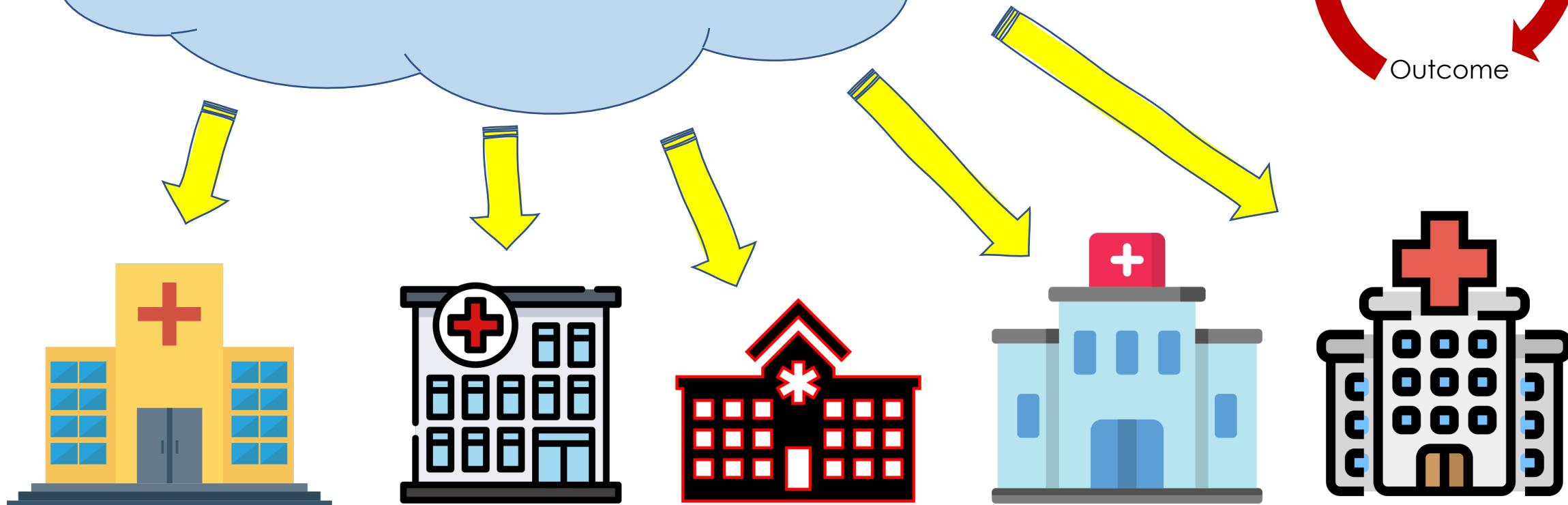
Infinite possibilities even within a specific domain



Case 5: Clinical Decision Making (Transferring patients)



Age, sex, injury severity index,
urgency, bed availability,
waiting time, and many
more...



Case 5: Questions to Consider

- Feature-availability at the initial presentation at the hospital/facility/emergency to determine the transfer facility
- What kind of model/models
- How to train the models
- How to deploy the models



This is a pretty complex problem

Case 6: Increasing Customer/Brand Engagement

- Directly related to customer retention
- Increased sales
- Increased revenue



Case 6 : Questions to Consider

- What medium or what features (factors) generate highest engagement

Case 7: Optimum Product-Bundle Recommendation

Customer Group A	Customer Group B	Customer Group C	Customer Group D
Product A		Product A	
Product B	Product B		Product B
Product C	Product C		

- The company makes more profit by selling more products
- Want to recommend product to customers that help saving them more (at the same time making more profit)
- Goal is to recommend product bundle to give them maximum saving

Case 8: Impact Evaluation + Recommendation

Client Group A	Client Group B
Received intervention "X"	Did not receive intervention "X"
<p>X could be any intervention</p> <ul style="list-style-type: none"> - e.g., telephone counseling on healthy eating - Only customers with this service added to their plans receives the phone counseling 	

GOAL:

- First, Assess the intervention (e.g., reduced cost, increased saving for customers)
- Second, if its impact was favorable, whether to recommend it to Group B?

- Observational data
 - Matching
 - Propensity
 - CEM
 - Regression
- Validation
 - How do you know your results are valid
- Recommend?





Summary

- Examples were progressively more complex
- Where do statisticians/data scientists fit?
- Where do machine learning engineers fit?

Credits

- Some Icons are made by Freepik
<https://www.flaticon.com/authors/freepik>
- Data Lake Image Source:
<https://blog.equinix.com/blog/2016/11/10/why-companies-are-jumping-into-data-lakes/>
- Stepped wedge trial schematic picture from
<https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0760-6>
- Pipeline picture from
<https://images.app.goo.gl/aMJMxDqTfETDESzM6>

- Apache Spark logo image from
<https://images.app.goo.gl/moGwHgHrAYfPiXjn7>