# Surgical Operations Quality Improvement

##### Group Members:

1). Joann Vuong [joann.vuong@utah.edu](mailto:joann.vuong@utah.edu) u0760934

2). Jeff Thornhill [jeffrey.thornhill@utah.edu](mailto:jeffrey.thornhill@utah.edu) u1328998

3). Theresa Aguilar [theresa.aguilar@utah.edu](mailto:theresa.aguilar@utah.edu) u1329017

##### GitHub Repository:

1). <https://github.com/tere646/Surgical-Data-Wrangling>

#### **Background and Motivation:**

#### We found surgical quality measures to be a particularly interesting subject due to the myriad of implications involved. Each of us have either had surgery ourselves or know someone has had surgery. Although we have not personally experienced post-operative complications, complications happen more often than what we would like. These complications can range from leaving surgical equipment in someone to infections during the recovery phase. It is estimated that 1 in 10 people experience an adverse event with a high volume of them being surgical in nature [2]. As the number of surgeries per year increase, there is literature finding that about 30% of patients who receive a procedure will experience a post-operative complication [3]. These post-surgery complications are costly to a health care system and increase the cost of healthcare. Fortunately, this is also an actionable area where health care cost can be reduced.

#### The COVID pandemic highlighted the importance of implementing known best practices like hand washing to reduce transmission ~~of~~ and contraction of disease The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) is taking a similar approach for surgical procedures by working with hospitals to develop best practices with the objectives of preventing surgical complications, saving lives, and reducing healthcare costs. As a group, we are excited to gain a deeper understanding of the measures and objectives the NSQIP have outlined. We believe there are simple best practices out there applicable to surgeries that would have a big impact on things such as reducing post-surgery sepsis.

#### **Project Objectives:**

While surgery is sometimes a necessity in repairing the body from sustained injuries or unexpected bodily ailments like an inflamed appendix, surgery brings about several potential risks and complications. Every surgery comes with a certain degree of risk. After reviewing the various NSQIP measures, there is a big emphasis in several different areas and how the patient is doing overall within the first 30 days post-surgery. First and foremost, preventing death due to surgical complications is at the top of the list. This can happen when a patient has complications like a heart attack, sepsis, pneumonia, and breathing impairment, among other things.

The primary question we are trying to answer is whether a hospital has the data necessary to participate in the NSQIP program. By determining the feasibility of a health systems participation we will be answering subsequent questions like:

* + What kind of time and resource would be required?
  + What are the necessary data elements needed to report on the identified NSQIP measures.
  + How is the hospital doing based on the NSQIP measure definitions.
  + Are there custom data elements that would be beneficial to incorporate into an EMR workflow that would make capturing measure performance easier.

As a team we are hoping to gain a better understanding of the overall NSQIP program because the healthcare landscape continues to shift towards a quality of care reimbursement model. NSQIP is one of several quality programs in the industry but they all contain technical specification documents for calculating their measures. This project will provide practical experience in interpretating requirements and translating them into a useable dataset. Additionally, we would like to develop our data wrangling skills when it comes to cleaning up erroneous data, converting data into a useable format, and grouping data into categories when necessary. The final thing we are hoping to learn through this project is displaying data in a concise and consumable way. Data cleaning is a necessary step but if the data analysis and insights are not presented in way that is understandable for end users, then the story that is trying to be told can be misinterpreted or ineffective.

Readmissions are common after surgeries and cost hospitals a substantial amount of money so it is important to understand which surgeries have a high likelihood of patients coming back as well as who so that the appropriate post-operative processes can be put into place. One study found surgery for a lower extremity vascular bypass to be as high as 14.9 percent [Need to site or delete]. Each hospital will vary on where they are seeing readmissions but the data collected by NSQIP and the subsequent analysis by our team will be beneficial in highlighting these high readmission rates by procedure. This will provide the hospital with a starting point of where to focus that could have an immediate impact in reducing cost and improving care.

#### 

#### **Data:** We will be collecting most of our data through the Centers for Medicare & Medicaid Services (CMS) website. The data provided by CMS provides a robust set of metadata for us to analyze. The data provided by CMS are distributed as CSV files, as such we will mostly be working with CSV files for this project. Once the data is gathered and cleaned, we will either consolidate the CSV files into few files or load them into a database. The link(s) for our data sources are provided below:

1). [https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use- Files/SynPUFs/DE\_Syn\_PUF](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF)

2). <https://www.facs.org/quality-programs/acs-nsqip>

#### **Data Processing:** We expect to do some level of data cleanup during this project. Since we are getting the files from CMS, we anticipate that there has already been some level of data quality checks and clean up that has been done. However, we did notice that some columns are in scientific notation which will require us to do some data cleaning in order to get the values in their desired format for analysis or data manipulation tasks. We anticipate that there are columns in the file that may not be in the correct datatype for aggregating the things we want to quantify. We are going to be doing counts by different categories (i.e. gender, age, etc.) and deriving columns like “Number of days between surgery and readmission” . These quantities will require numeric datatypes to improve processing performance and aggregation. Additionally, to increase our understanding of the data, we may modify column names that have lesser known acronyms so that the data set(s) is easier to comprehend by a broader audience. All of these components of mapping, transforming, and cleaning we anticipate doing during the data processing phase in order to set ourselves up for better data analysis.

#### **Design:**

Before displaying the actual results of the data, it is important to create a workflow diagram that maps out the different inclusions and exclusions for the various measures in order to transparently show how the dataset(s) was developed. These are important steps to take so that users can more fully understand the visualizations that they are looking at. For these measures, it will be important to display the data through different demographic aggregations like gender, race, age groupings, etc. This will allow us to understand who is experiencing surgical complications and if there are any populations that are more susceptible to adverse effects than others. This data will be displayed using bar charts to show frequencies in a side by side comparison. Histograms can be used to show a frequency of days that have passed before a patient came back. Or, depending on how many categories there are, a pie chart could be considered but those quickly become distorted as the number of slices increases.

A big benefit of NSQIP is their ability to do benchmarking. If the right data elements and amount of data is available, we plan on comparing across facilities and doing our own site comparisons. These graphs are important as patients decide on where they want to have their surgical procedures completed. This can be shown using stacked graphs over time or site ratios. Or, a combo bar and line chart will be beneficial in this scenario. In order to create these different visualizations, we plan on using one of the many free software options available like PowerBI, Qlikview, R, RShiny, or Python. There are pros and cons to each of these products but each of them have the necessary visualization components in order to properly display the data. The two main influences in our visualization tool of choice will be, team member experience and portability across different laptops.

As an alternative design, we can use Excel and the pivot tables it provides to create charts and tables. Sometimes, numbers in a tabular format are just as effective as bar charts. Plus, the benefit of Excel is that it is very a common tool and has familiarity with a wide audience.

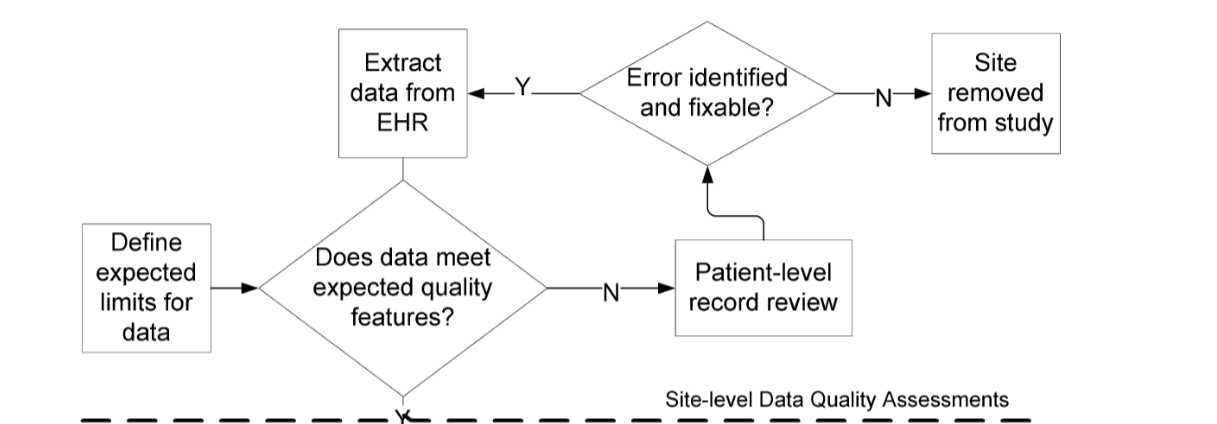
#### **Must-Have Features:** List the features without which you would consider your project to be a failure.

Measure definitions are vital to the success of the project because users need to know conceptually what is being measured, why it is necessary, and how to actually calculate each measure. Without a clear and thorough understanding of the inclusion and exclusion requirements the project is bound to deviate from the desired NSQIP calculations and therefore prevent an apples to apples comparison of other hospitals in the NSQIP program.

A second requirement to be successful in our project will be having the relevant data elements required for calculating the defined NSQIP measures. This will include things like birthdays, admission/encounter dates, diagnosis information, etc. Depending on the data size, a cloud option or SQL database may be required for optimal processing. The final feature of this project is the presentation of the data. We believe that if the data analysis and visualizations that are completed in this project cannot be communicated in an effective way, the data will be in-actionable and actionable change will not be able to take place. The features mentioned in the design portion of the process book are also essential.

#### **Optional Features:** List the features which you consider to be nice to have, but not critical.

Ideally we would provide a clear process model of our data quality assessment procedure allowing readers to more easily understand the workflow. A side by side comparison of pre-cleaned data and post cleaned data will demonstrate the value of data clean up. A visual of our process model connecting to the NSQIP process model would also be nice to put our work in perspective. Perhaps a screen shot of the NSQIP data entry portal in order to demonstrate how data clean up makes for smoother data entry.

  
*Above is an example process model from Kahn et al.*

A side-by-side comparison of pre-cleaned data and post-cleaned data will demonstrate the importance of data clean up. Perhaps including a screen shot of the NSQIP data entry portal in order to demonstrate how the clean data makes for smoother data entry. Also creation a list of necessary data changes for future data analysis on these datasets and/or be able to reproduce our results at a future date.

Lastly, additional location data features could provide further research from a benchmarking perspective, and identifying possible access of care issues for patients. Analyzing location data may provide insights into how two or more places are connected or correlated.

#### **Project Schedule:** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

**Week 1:**

1). The main objective for the first week is to work on the project proposal. Each of the sections of this document will be divided equally amongst team members.

2). The secondary objective is to start looking into what datasets are available to work with around NSQIP measures.

**Week 2:**

1). The main objective for week 2 is to review the project proposal as a group to ensure that we have answered each of the questions and collaborate together on any outstanding questions.

2). We also want to have all of our datasets identified as to where we are going to get them from and downloaded in preparation for the next step in the process: Data Cleaning.

**Week 3:**

1). Week 3 is all about data cleaning. This is the primary objective for the week as data cleaning can take a lot of time and will be conduct by everyone. The different datasets that have been identified from the previous week will be divided evenly across team members. During this week, we will be evaluating which columns are needed, adjusting any columns that may not be in a usable datatype, and derive any additional columns that may be helpful in the future for visualizations or analysis.

2). The secondary objective will be to determine our loading strategy. Are we going to consolidate the different CSV files into one giant file or upload them into a cloud solution like Oracle or use a SQL database, etc.

**Week 4:**

1). During week 4, the team will meet together in order to start the data loading process. At this point, we will know where we are going to load the data and will begin using either the built in tools of the database clouds or an open source tool like Pentaho.

**Week 5:**

1). The primary objective will be data analysis. Once the data has been loaded and combined from the previous week, we will spend time analyzing the data. The coding language we use will be influenced by team member experience and where the data is located. If it ends up in a database, analysis will be conducted mostly using SQL but if it ends up in a series of aggregated CSV files, either Python or R will be used.

2). The secondary objective during this week is to begin building out the specified visualizations in either PowerBI or Excel. Every team member will work on each of these weekly objectives but the work will be divided by the measures that need to be calculated. This will allow each team member to get practical experience in each of the project steps that will be transferrable in our future jobs.

#### **Week 6:**

1). Data visualizations will be completed in week 6. Each person will have worked on different pieces so we will spend time consolidating all of the pieces into one dashboard and file so that everything is located in one place.

2). The second objective during this week is to work on creating and developing all of the project deliverables in preparation for our final presentation.

**Week 7:**

1). Week 7 has us finishing up our project and practicing our presentation. This week will focus primarily on polishing up any outstanding project deliverables and determining who is going to be presenting what, how, etc.

#### **References:**

1). https://media.jamanetwork.com/news-item/hospital-readmissions-after-surgery-associated-mostly-with-complications-related-to-surgical-procedure/

2). <https://pubmed.ncbi.nlm.nih.gov/18519629/>

3). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3637983/