The proposal document should address the following points. Use these points as headers in your document.

* **Basic Info.** The project title, your names, e-mail addresses, UIDs, a link to the project repository.

Title:

Team Members:

* Matt Alexander ([matthew.alexander@hsc.utah.edu](mailto:matthew.alexander@hsc.utah.edu); u6013267)
* Jason Funderburk ([jason.funderburk@utah.edu;](mailto:jason.funderburk@utah.edu) u1377566)
* Kelsey Nicholson ([u1315045@utah.edu](mailto:u1315045@utah.edu); u1315045)
* Matthew Poulsen ([matthew.poulsen@utah.edu](mailto:matthew.poulsen@utah.edu); u0546346)
* Elizabeth Sloss ([liz.sloss@nurs.utah.edu](mailto:liz.sloss@nurs.utah.edu); u6038479)
* John Stanley ([u0873475@umail.utah.edu](mailto:u0873475@umail.utah.edu); u0873475)

Link to Project Repository:

https://github.com/Mdpoulsen/BMI\_6016\_Group\_Project\_NSQIP.git

* **Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Our group has a well-balanced composition of clinicians, researchers, and data scientists. NSQIP represented a subject that overlapped each of the disciplines in an interesting way making it the project of choice. For some, it represented an opportunity to gain a deeper learning about the NSQIP process, how to objectively view the data, learn the limitations of NSQIP and CMS data, and gain insight on how to meaningful impact these measures. For others, the CMS dataset represents an opportunity to learn about data that is interesting but not part of their daily work.

* **Project Objectives.** Provide the primary questions you are trying to answer with your data. What would you like to learn and accomplish? List the benefits of how the data could be useful.

For this project, our team is acting as a health data company assisting with evidence-based improvement of surgical care quality for hospital systems. The National Surgical Quality Improvement Program (NSQIP) is an outcomes-based quality improvement program through which we identified 5 specific critical quality metrics relevant to surgical care and surgical care quality. These quality measures included prevalence of surgical site infections (SSIs), patient readmissions following surgical procedures, reoperation within 72hrs of initial procedure, length of post-operative medical-surgical unit stay (LOS), and instances of post-operative deep-vein thromboses (DVT’s). These quality metrics and data will provide a foundation for which evidence-based quality improvement initiatives can be built for each hospital system.

* **Data.** From where and how are you collecting your data? If appropriate, provide a link to your data sources.

For this project, we will use data from Medicare Claims Synthetic Public Use Files   
(SynPUFs). Medicare Claims SynPUFs were created by the Centers for Medicare & Medicaid Services (CMS) from claims data and are available for free download. CMS creates SynPUFs by randomly sampling 5% of Medicare beneficiaries and synthesizing the data. Specifically, we will utilize the 2008-2010 data which was collected from Medicare beneficiary claims between 2008 and 2010. The data file includes the following information for individual beneficiaries: demographic, clinical, and financial/economic. The following link can be used to access SynPUFs: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DESample01>

Further, a user manual for Medicare Claims SynPUFs can be found here: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_DUG.pdf>.

Summaries of variables within the dataset can be found here:

<https://www.cms.gov/files/document/de-10-codebook.pdf-0>

ICD-9-CM codes for each year can be found here:

<https://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/codes>

* **Data Processing.** Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

For data processing, we plan to import all inpatient claims data from SynPUFs into python data frames. We will also import diagnostic and procedural ICD-9-CM codes as data frames. Since multiple versions of ICD-9-CM datasets were put into effect during the 2008-2010 time period, we will assess differences between the ICD-9-CM during the years and ensure appropriate use of the different versions. For testing purposes, the separate sample files of inpatient/outpatient claims data may be kept separate, however we plan to eventually combine the imported data into a unified data frame, provided it is not too computationally expensive.

Preparing data for our quality measures requires identifying ICD-9-CM codes that match certain criteria, such as surgical site infections and surgical procedures. To match codes with criteria for data quality measures, we plan extend our ICD-9-CM data frame to include columns containing Boolean values to denote a non-match (0) or match (1) between the code and each of the predetermined criteria.

For almost all surgical quality measures, extracting the relevant data requires looking across multiple instances of a patient’s claims data. For this, we will group the data by patient unique identifiers. Patients with zero history of surgeries will be excluded from the data frame to reduce space and quicken computations.

Each of the 5 quality measures will be extracted using related methods. For prevalence of surgical site infections (SSIs), patient readmissions following surgical procedures, reoperation within 72hrs of initial procedure, and instances of post-operative deep-vein thromboses (DVT’s), dates connected with surgery will be compared to dates that fall within a given timeframe and match criteria for one of these quality measures. Counts of these instances and will be extracted, along with a count of surgeries across the entire dataset. For post-operative medical-surgical unit stay (LOS), which does not require patient grouping, we will extract counts for each length of stay following surgery.

* **Design.** How will you display your data? Provide some general ideas that you have for the design. Develop **one alternative prototype design for your data**.. Describe your designs and justify your choices of visual encodings.

Our project is focused on the feasibility of using CMS claims data to serve as entry data for NSQIP. In a real-world presentation, our goal would be to demonstrate that CMS data can be appropriately prepared for entry into NSQIP thus saving valuable time of the NSQIP coordinator. To demonstrate the feasibility of the project, we will show the measures as obtained from the CMS data and compare with national benchmarks. Given that our audience would likely be hospital administrators and given that we would be comparing items (each individual measure vs benchmark) over time (each year in the synthetic CMS data set), a simple approach of a column chart could visually demonstrate parity or superiority of the scrubbed CMS data to national NSQIP benchmarks.

An alternative more data intensive approach would be to utilize a small pivot table. This would allow us to reveal the data in more detail when grouped by year and then by measure. This approach would be more valuable when presenting to the NSQIP coordinator or the Quality, Safety, and Value Officer as they often have a greater appreciation for a more granular approach.

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

Analysis of data completeness with graphic representations. Identification of missigness variance among clinical groups. Summary statistics for centers and clinical services for the five critical quality metrics.

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

Interactive data visualization to display each of the quality metrics with dynamic visualization by selecting any number of medical centers, healthcare systems, clinical teams, etc.

* **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 of 2/7: Complete project proposal, create Github repo; Prepare for data description presentation

Week of 2/14: Meet w/ faculty, incorporate feedback into proposal and project plan

Week of 2/21: Work on data processing [initiate for 2 of 5 quality measures]

Week of 2/28: Work on data processing [continue to work on remaining quality measures]

Week of 3/7: Submit project update via Github; Receive and incorporate peer feedback

Week of 3/14: Continue to work on data processing

Week of 3/21: Submit project update via Github

Week of 3/28: Continue to work on data processing [finalize data processing for all quality measures]

Week of 4/4: Prepare for data wrangling intermediate work presentation

Week of 4/11: Data Wrangling Intermediate Work Presentation

Week of 4/18: Finalize project submission and prepare for final presentation

Week of 4/25: Present final project and submit supporting documentation

Week of 5/2: Submit final project

This proposal is the first part of your process book. As a ballpark number: your proposal should contain about 3-4 pages of text.

You will schedule a project review meeting with a staff member during regular lecture times of the week marked in the schedule. Make sure all of your team members are present at the meeting.

The proposal will be submitted by uploading it to your team’s GitHub repository.