SIADS 593: Milestone I

Team Project Proposal

version 2022.07.27.1.CT

**Instructions:** please make a ***copy*** of this template file (do not edit original).

## **Proposal Title:**

## Evaluating Performance in End-Stage Renal Disease (ESRD) Facilities and Clinician Groups: A Comparative Study of the ESRD Treatment Choices (ETC) Model

## 

## 1. Team members

Please list your team members (2-3 max).

* Iris Lin
* Kasra Afzali
* Michael Light

## 2. Project summary

Summarize your proposed project in a few sentences.

#### What is your proposed project and why are you proposing it?

#### What are the question(s) you want to answer, or goal you want to achieve?

| * The ESRD Treatment Choices (ETC) Model was designed to improve outcomes for patients with end-stage renal disease by encouraging home dialysis and kidney transplants. * This project aims to evaluate the performance of both ESRD facility aggregation groups and managing clinician groups using key performance metrics such as Performance Payment Adjustment (PPA), Modality Performance Score (MPS), home dialysis rate, and transplant rate. * By comparing these performance measures, we seek to understand differences in outcomes between facilities and clinicians and provide insights for optimizing care models and payment adjustments. |
| --- |

## 3. Datasets

#### Describe one primary dataset and at least one secondary dataset. If other secondary datasets will be used please describe them as well.

#### The proposed datasets should exhibit different features/columns and/or different access methods, e.g., \*.csv file, \*.json file, API retrieval, web scraping, etc. Different time periods, for example, with the same features/columns is not considered a different dataset. Remember, the focus of the project in this Milestone course is to give you the opportunity to practice your data manipulation skills, so feel free to challenge yourself.

#### If you're unsure if your data sets are "different enough" describe the datasets and request a review via the *#siads593\_[semester]\_001\_project* Slack channel.

#### **Please note:** all proposed datasets ***MUST*** be publicly available to all members of the class (students, instructors, course support personnel, etc.). Use of proprietary datasets for this project is ***not*** permitted.

## 3.1 Primary dataset description

Describe your primary dataset. How is the data collected and how will you access it? Please share what features in the dataset are relevant to your topic. At a minimum, include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

| **Datasets:**   * **Dataset 1: ESRD Facility Aggregation Group Performance**   + Source: Centers for Medicare & Medicaid Services (CMS)   + Features: Performance Payment Adjustment (PPA), Modality Performance Score (MPS), home dialysis rate, transplant rate, facility aggregation group information   + Quality: Comprehensive but may include regional biases due to geographic selection of ESRD facilities (30% of HRRs, excluding U.S. Territories)   + Size: 421 rows   + Format: csv   + Public   + Has some missing/supressed values   Dataset 1:  [End-Stage Renal Disease Facility Aggregation Group Performance | CMS Data](https://data.cms.gov/cms-innovation-center-programs/end-stage-renal-disease-treatment-choices-model/end-stage-renal-disease-facility-aggregation-group-performance/data) |
| --- |

## 3.2 Secondary dataset(s) description

Describe your secondary dataset(s). How is the data collected and how will you access it? Please share what features in the dataset(s) are relevant to your topic and describe the data types you’re expecting. At a minimum, for each secondary dataset include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

| [Please use this space for your response. You may expand or contract this box as needed.]   * **Dataset 2: Managing Clinician Aggregation Group Performance**   + Source: Centers for Medicare & Medicaid Services (CMS)   + Features: Performance Payment Adjustment (PPA), Modality Performance Score (MPS), home dialysis rate, transplant rate, clinician aggregation group information   + Quality: Similar geographic selection as Dataset 1, focusing on HRRs and Maryland-specific ZIP codes   + Size: 900 rows   + Format: csv   + Public   + Has some missing/supressed values   Dataset 2:  [Managing Clinician Aggregation Group Performance | CMS Data](https://data.cms.gov/cms-innovation-center-programs/end-stage-renal-disease-treatment-choices-model/managing-clinician-aggregation-group-performance/data) |
| --- |

## 3.3 [ ] Affirm: datasets are public.

Please write YES in the above box to confirm that your primary and secondary datasets are accessible and available to your classmates and the instructional team.

## 4. Cleaning and manipulation

Describe how you will need to manipulate your datasets: how will you handle missing or anomalous data? How will you join your primary and secondary datasets? What cleaning and manipulation challenges, if any, do you anticipate?

| **Cleaning and Manipulation:**   * **Data Collection:**   + Gather both datasets from CMS repositories for the same time period to ensure comparability.   + Use crosswalk files to map aggregation groups to individual ESRD facilities and clinicians.     - [Supplemental Data - Dartmouth Atlas DATA](https://data.dartmouthatlas.org/supplemental/#crosswalks) * **Cleaning Steps:**   + Handle missing or incomplete performance scores by using appropriate imputation techniques or removal where necessary.   + Standardize features like PPA and MPS across both datasets for uniform comparison.   + Remove duplicate aggregation groups or regions that appear in both datasets. * **Preparation Techniques:**   + Create derived metrics such as relative performance improvements (e.g., change in home dialysis rate).   + Merge facility and clinician data on common dimensions like geographic region and aggregation group. |
| --- |

## 5. Analysis

Describe any analyses you plan to undertake. For each, please give the technique or approach and briefly explain what you expect to learn from it.

| [Please use this space for your response. You may expand or contract this box as needed.]  **Questions & Objectives:**   * How do performance metrics differ between ESRD facilities and managing clinicians in similar geographic areas? * What factors contribute most to higher home dialysis and transplant rates? * Can payment adjustments (PPA) be optimized based on performance differences between facilities and clinicians?   **Formulations:**   * Comparative analysis using regression models to explore relationships between home dialysis rates, transplant rates, and payment adjustments. * Correlation analysis between MPS, home dialysis rates, and geographic location. * Time-series analysis to track performance changes over time and across different HRRs. |
| --- |

## 6. Visualizations

Describe in 1-3 sentences at least **two** data visualizations that you plan to create. Include the chart type (e.g. bar chart, scatterplot, SPLOM, etc.) as well as the variables (features) you intend to plot.

| [Please use this space for your response. You may expand or contract this box as needed.]  **Graphs & Visualizations:**   * Bar charts comparing PPA and MPS across facilities and clinicians. * Scatter plots illustrating the relationship between home dialysis rates and performance scores. * Regional heatmaps displaying performance disparities across HRRs. |
| --- |

## 

## 7. Ethical considerations

Does your choice of data raise any ethical issues? If so, briefly describe the concern and how you plan to mitigate it.

| * **Patient Privacy:**   + Although the data is aggregated, ensure no individual patient information is identifiable. * **Geographic and Demographic Bias:**   + The geographic selection of certain HRRs could lead to bias in results, potentially overlooking underserved populations.   + Plan to mitigate: Acknowledge geographic limitations and adjust analyses to account for regional disparities. * **Healthcare Equity:**   + Differences in access to home dialysis or transplant options may disproportionately affect certain groups.   + Plan to mitigate: Examine disparities across regions and demographics, ensuring findings are interpreted with equity in mind. |
| --- |

## 8. Contributions

Indicate the contribution that each team member will make to the project.

| [Please use this space for your response. You may expand or contract this box as needed.] |
| --- |

### 

## Changelog

(2022.07.27.1.CT) Update for 593

(2021.07.24.1.AW) Adjust title, number sections, simplify section headings, edit text