

An Exploration of Voting Drivers Behind Dialysis Regulation in California

1 Project Motivation

1.1 Background on Dialysis and Proposition 29

Dialysis is a life-sustaining treatment that purifies blood for patients with chronic kidney disease, effectively serving as an artificial kidney. In recent years, California has seen multiple attempts to increase regulations on dialysis clinics, with Proposition 29 in 2022 being the latest in a series that included Proposition 8 in 2018 and Proposition 23 in 2020. All three propositions failed to pass.

Proposition 29, known as the Dialysis Clinic Requirements Initiative, sought to implement stricter regulations on staffing and operations for California's approximately 600 dialysis clinics—an industry valued at an estimated \$3.5 billion. A key provision of the proposed regulation mandated the presence of a physician or licensed practitioner during all treatment hours, a requirement that would have significantly increased operational costs for each clinic by several hundred thousand dollars annually.

1.2 Debate Surrounding Dialysis Regulations

- **Proponents** argue that increased regulations improve patient safety and quality of care.
- **Opponents** contend that the increase in healthcare costs is unwarranted and would limit care coverage by overwhelming facilities with costs and potentially **forcing them to close**.

1.3 Project Scope and Significance

Our project explores trends in dialysis clinic access, quality of care, and ballot results in the state of California in recent years. Using publicly available data from the Center for Medicare and Medicaid Services and aggregated election results from California's Secretary of State, we've analyzed associations between dialysis care and voting behaviors, specifically those related to recent statewide ballot initiatives designed to regulate California's multibillion-dollar dialysis industry—including 2022's Proposition 29, which failed to pass by a large margin.

1.4 Relevance and Novelty

- This project uses a novel approach to study an area of public interest relevant not just to California but the entire country.
- To our knowledge, this is the first project of its kind to explore the possible association between voting patterns and the quality of dialysis care received by patients at the facility level.
- The majority of dialysis treatments in California **are covered by Medicare**, making the Medicare and Medicaid datasets particularly relevant to our analysis.

1.5 Broader Context

Investigative reporters, patient advocacy groups, and labor organizations have spent significant resources over the past decade to raise public awareness of the dialysis industry and its need for regulation. Our project contributes to this ongoing discussion by providing data-driven insights into the relationship between dialysis care quality and voting behavior.

2 Project Research Questions

Our research questions are divided into two categories: primary and secondary. The primary question serves as the main focus of our analysis, while the secondary question provides additional insights through further investigation of the data.

2.1 Primary Research Question

Is the quality of care at dialysis facilities correlated with voting in favor of or against dialysis industry regulation?

2.1.1 Key Assumptions

1. The relationship between Quality of Care and Voting Behavior is not confounded.
2. A vote in favor of any of the three propositions (Prop 8 in 2018, Prop 23 in 2020, Prop 29 in 2022) can be interpreted as support for dialysis industry regulation broadly speaking.

2.1.2 Quality of Care Metrics

To test this relationship under the outlined assumptions, we approximate quality of care using the following metrics, measured at the facility level:

- Five-star rating
- Patient experience rating
- Facility mortality rate
- Number of available dialysis stations
- Staff rating
- Hospital readmission categorization (Worse than Expected, As Expected, Better than Expected)
- Profit/non-profit designation
- Parent company affiliation/independence

2.1.3 Facility Categorization

Observations of Quality of Care metrics in our data are categorized by:

- Year
- County
- City
- Profit/Non-profit designation
- Parent company affiliation/independence

2.2 Secondary Research Questions

1. What is the geographic coverage of dialysis facilities in California?

3 Data Sources

3.1 Primary Data Sources

3.1.1 CMS Quarterly Dialysis Facility Compare Dataset

- **Type:** .zip & .xlsx
- **Years:** 2017-2024
- **Size:** 5,684 rows, 173 columns

- [Link: CMS Dialysis Facilities Data](#)

Key features: - Star ratings for facilities - Patient experience metrics - Quality of care metrics

Insights provided: - Patient satisfaction - Clinical outcomes - Doctor-patient communication - Hospitalization rates - Treatment effectiveness

Rating calculation: - Patient experience: bi-annual surveys - Facility ratings based on: - Unplanned hospital readmissions - Total and expected transfusions - Ratio of deaths to expected deaths - Waste removal efficiency

3.1.2 CA Secretary of State's Statement of Vote

- **Type:** .xlsx
- **Years:** 2018, 2020, 2022
- **Size:** 7,122 rows, 7 columns
- [Link: CA SOS Statewide Election Results](#)

Elections: November 2022, 2020, and 2018 **Focus:** Propositions on dialysis clinic requirements

Geographic levels: - Counties - Sub-counties: - Congressional districts - State senate districts - State assembly districts - Cities

3.2 Secondary Data Source

3.2.1 CA Health and Human Services Specialty Care Clinic Data

- **Type:** .xlsx
- **Years:** 2013-2023
- **Size:** 6,605 rows, 143 columns
- [Link: CHHS Specialty Care Clinic Data](#)

Purpose: Supplement CMS dataset with geographic data **Additional features:** - Senate district - Congressional district - Latitude and longitude

3.3 Data Integration and Analysis Potential

- Multiple geographic levels for varied scale analysis
- Clinical (CMS) and voting (SOS) data combination enables correlation exploration
- Enhanced spatial analysis with supplementary geographic data

4 Data Manipulation Methods

Our workflow was broken down into four stages:

1. Data Collection
2. Data Preparation
3. Exploratory Data Analysis
4. Statistical Analysis

4.1 Data Collection and Preparation

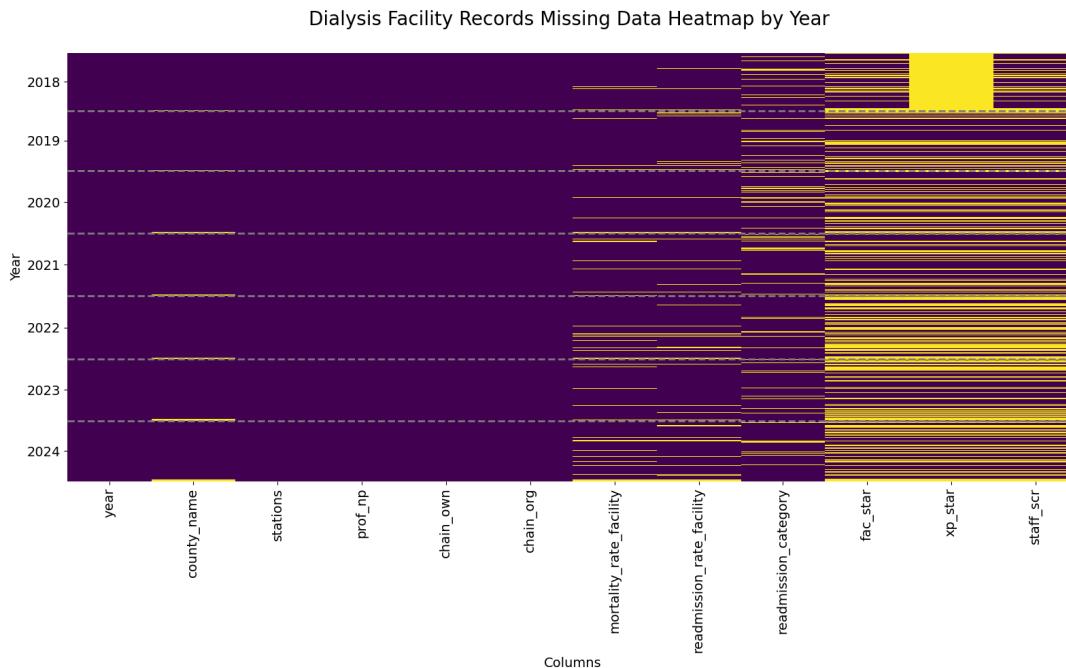
4.1.1 CMS Dialysis Facility Dataset

4.1.1.1 Organization and Import

- Dataset structure: .zip files (one per year), containing multiple Excel files
- Focus: Excel files relevant to facility general information, ratings, and patient survey results
- Import result: Two separate parquet files at the facility level
 1. Patient survey responses
 2. Facility ratings and measurements

4.1.1.2 Challenges and Solutions

1. Inconsistent File Naming Conventions
 - Issue: 2021 files named differently (e.g., patient survey data file named ‘59mq-zhts’ instead of including ‘ICH’ for In-Center Hemodialysis CAHPS Survey)
 - Solution: Created a list of exact file names for selection, rather than using pattern matching
2. Missing Data
 - Expected missing data: Survey non-responses
 - Unexpected missing data: Administrative errors (e.g., missing columns in recent ICHPS raw data files, including patient hospital readmission categorization and overall patient experience ratings)
 - Solution for specific cases: Simple imputation during analysis (e.g., substituting 2018 ‘nan’ values with 2019 values at the facility level)



4.1.2 SOS Ballot Data

4.1.2.1 Import and Selection

- Data imported via URL for each relevant proposition year (2018, 2020, 2022)
- Selected columns containing ‘Kidney’ or ‘Dialysis’ for analysis
- Geographic column manipulation:
 - Renamed columns
 - Backfilled rows to address multi-level index (sub-counties under counties)
- Final output: Single ballot data parquet file
 - Includes year column
 - Count and sub-county vote counts for each Dialysis Requirements Initiative proposition

4.1.2.2 Challenges and Solutions

- Inconsistent naming conventions across years
 - 2020 and 2022: ‘County Supervisorial’
 - 2018: ‘Supervisorial District’
- Solution: Standardized naming across all years

4.1.3 CHHS Specialty Care Clinic Complete Data Set

4.1.3.1 Import and Alignment

- Downloaded Excel files for 2013 through 2023 (one per year)
- Main challenge: Aligning pre-2018 data with 2018-forward structure
- Process:
 1. Separated data into two dataframes: 2013-2017 and 2018-2023
 2. Used CHHS mapping dictionary to rename 2013-2017 columns
 3. Ensured consistent data types across both dataframes
 4. Merged dataframes using outer join on common columns
 5. Dropped rows with missing FAC_NO (facility data)

4.1.4 Data Merging and Standardization

- Standardized data types and column names across all datasets
- Merged datasets:
 1. CMS facility rating dataset with CMS patient survey dataset
 2. Filtered CHHS dataset (dialysis clinics only) with merged CMS data
 3. Reshaped CMS and CHHS data by geographic level
 4. Merged geographic-level data with SOS Ballot Measures dataset

4.1.5 Final Output

- Two parquet files:
 1. Data aggregated at city level
 2. Data aggregated at assembly district level

4.2 Additional Details on Data Manipulation

4.2.1 CMS Dialysis Facility Dataset

The dataset was organized in .zip files, one per year, each containing several Excel files grouped by variables. We focused on files relevant to facility general information, ratings, and patient survey results. The data was split into two separate parquet files at the facility level: one for patient survey responses and another for facility ratings and measurements.

4.2.2 SOS Ballot Data

The data was imported by URL for each relevant proposition year (2018, 2020, and 2022). We selected columns containing ‘Kidney’ or ‘Dialysis’ for analysis. Geographic column manipulation included renaming columns and backfilling rows to address the multi-level index with sub-counties falling under their respective counties. The cleaned data for each year was merged into one final ballot data parquet with a column specifying the year and count and sub-county vote counts for each Dialysis Requirements Initiative proposition.

4.2.3 CHHS Specialty Care Clinic Complete Data Set

Excel files were downloaded for 2013 through 2023, one file per year. The main manipulations were performed on files from years before 2018 to align them with the structure and naming convention of the

files from 2018 onward. We used a mapping dictionary created by CHHS to map pre-2018 and post-2018 data columns, ensuring consistent column names and data types across all years.

4.2.4 Data Merging Process

To merge the datasets on time period, geography, and facility names, we standardized data types and column names. The process involved:

1. Merging the CMS facility rating dataset with the CMS patient survey dataset
2. Filtering the CHHS dataset to show only dialysis clinic-related data
3. Merging the filtered CHHS data with the combined CMS data
4. Reshaping the CMS and CHHS data by geographic level from the original facility-level information
5. Merging these aggregations with the SOS Ballot Measures dataset

This comprehensive data manipulation process resulted in two final parquet files: one with data aggregated at the city level and another at the assembly district level.

5 Analysis and Insights

This project employed a Bayesian approach to investigate the relationship between dialysis facility quality metrics and voting patterns on dialysis-related propositions in California.

To maximize the granularity of our analysis, we chose to focus on city-level data, which provided more detailed vote counts compared to assembly district level data. We encoded voting outcomes as the percentage of “Yes” votes in favor of the propositions, allowing for a nuanced examination of support for dialysis industry regulation across different localities.

Given the limitations and challenges detailed below, we were able to gain some insights into our primary research question: **Is the quality of care at dialysis facilities correlated with voting in favor of or against dialysis industry regulation?**

We were also able to use visual analysis of the un-modeled data to gain some insights into the geographic coverage of dialysis facilities in the state.



Dialysis Stations per capita 2022

5.1 Analysis Steps

5.1.1 Data Preparation

Before modeling and analysis, we underwent several data preparation steps. These steps were performed during the analysis stage in the interest of transparency. Steps taken included:

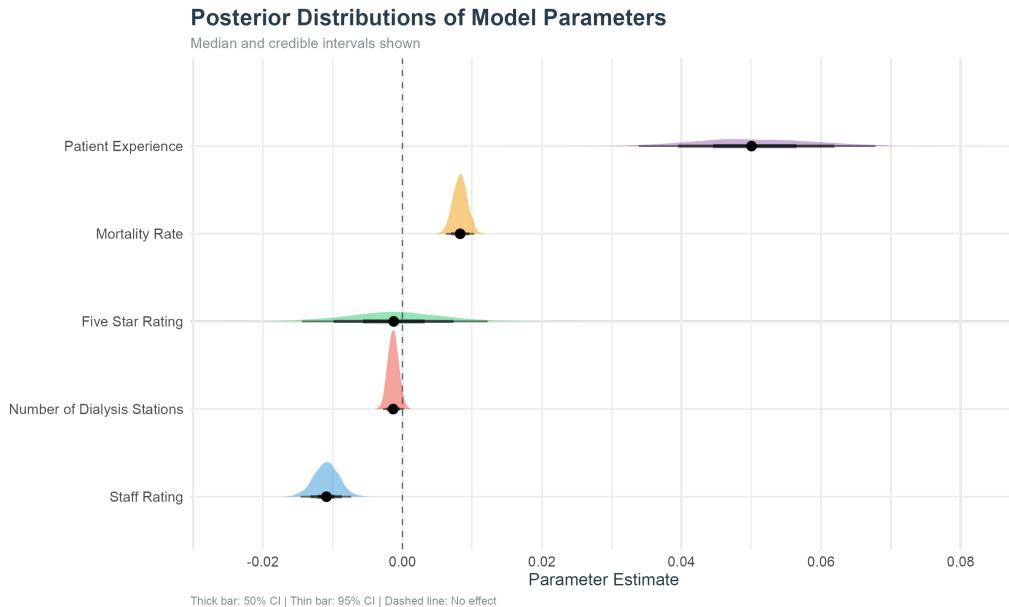
1. Imputing missing 2018 values using 2019 data for select variables.
2. Converting datatypes.
3. Calculating vote percentages in favor of regulation for each facility's city.
4. Filtering data to include years 2018, 2020, and 2022.
5. Aggregating data at the facility level, summarizing vote outcomes and facility characteristics.
6. Removing rows with missing values to perform a complete case analysis.

5.1.2 Model Construction

We constructed a Bayesian multilevel model using the `brms` package in R. This model allowed us to account for the hierarchical nature of our data (facilities nested within cities and counties) while examining the relationship between facility quality metrics and voting outcomes.

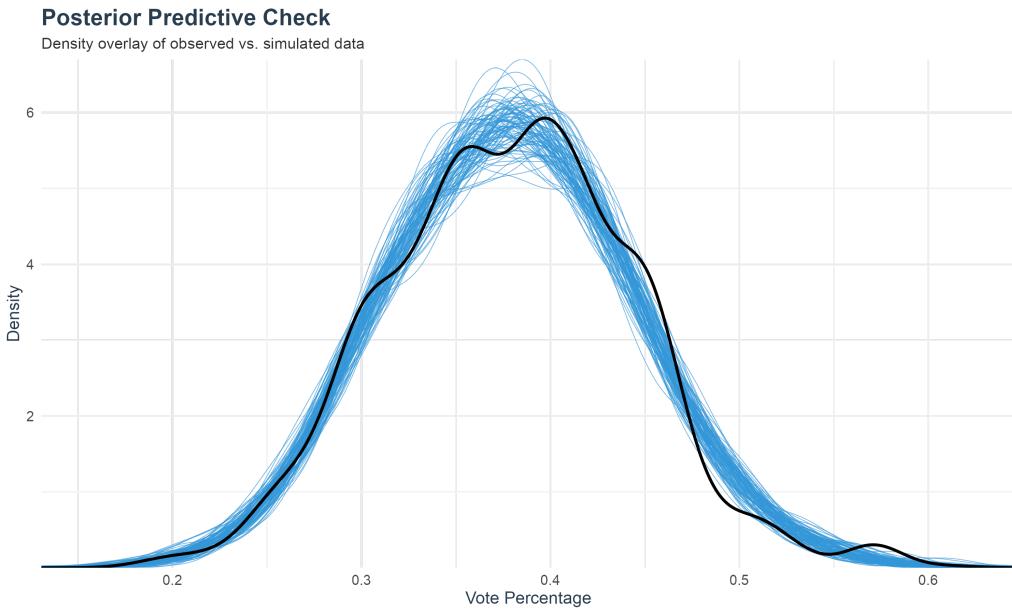
5.1.3 Posterior Predictive Checks:

We performed posterior predictive checks to assess model fit and explore relationships between key variables and voting outcomes. The following visualization provides a comprehensive overview of the model parameters and their effects on the predicted vote percentage in favor of regulation.



Posterior Predictive Check for Chain Organizations

To assess the overall performance of our model in capturing the relationship between voting behavior and quality of care metrics, we conducted a posterior predictive check for the entire model. This check allows us to compare the observed data with simulated data from the fitted model, providing insight into how well the model represents the underlying patterns in our dataset.

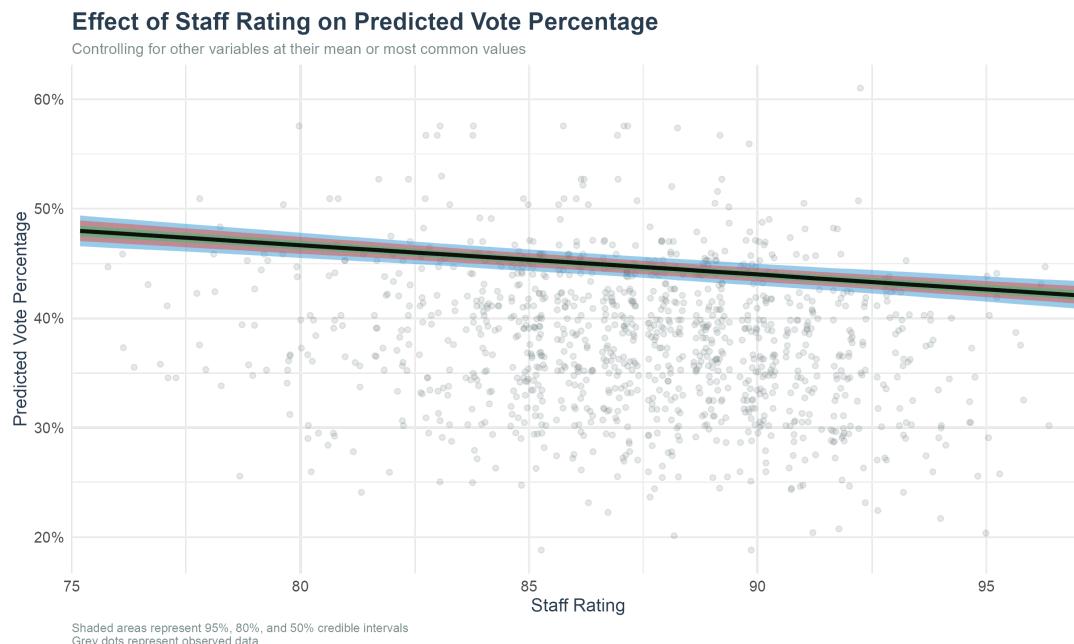


Posterior Predictive Check for Entire Model

5.2 Insights

5.2.1 Staff Rating Impact:

Our analysis revealed a negative relationship between staff ratings and the percentage of votes in favor of dialysis regulation. This suggests that areas with lower-rated dialysis facility staff were more likely to support increased regulation. The effect of staff rating varied across counties, with some counties showing stronger negative relationships than others.

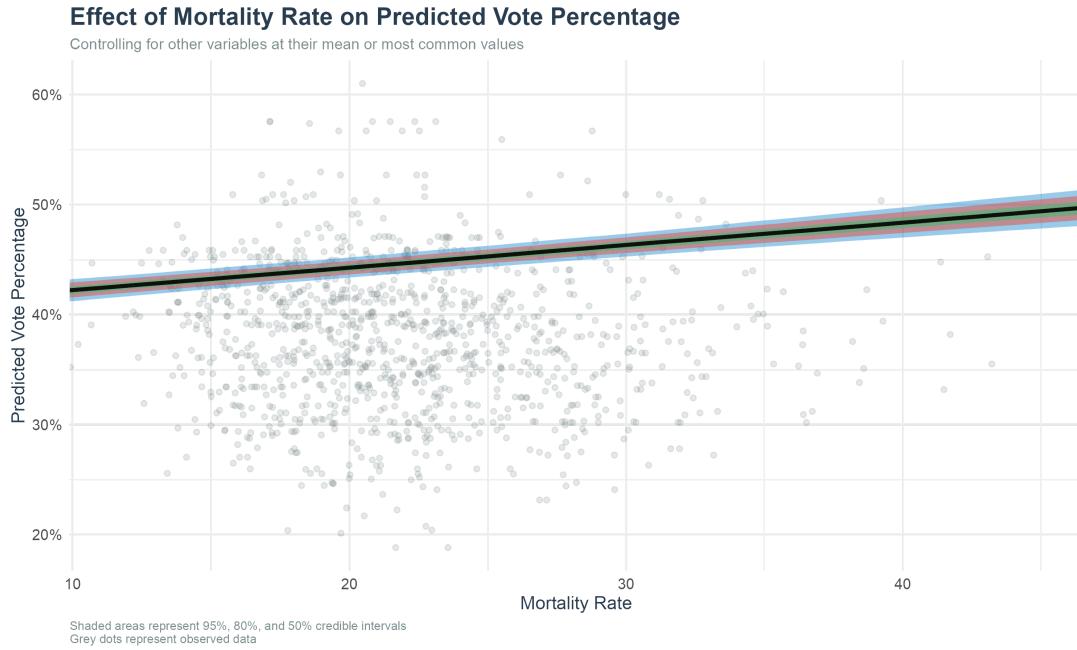


Effect of Staff Rating on Predicted Vote Percentage

5.2.2 Mortality Rate Influence:

We found a positive relationship between facility mortality rates and support for regulation. As mortality rates increased, the predicted vote percentage in favor of regulation also increased. This suggests that

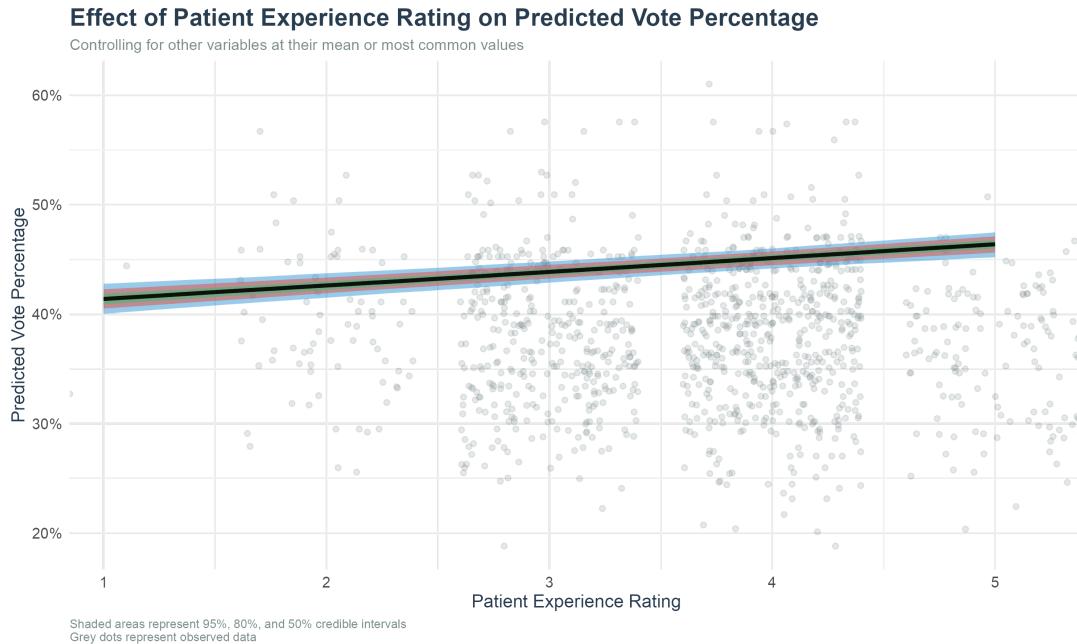
voters in areas with higher mortality rates at dialysis facilities were more likely to support increased oversight.



Effect of Mortality Rate on Predicted Vote Percentage

5.2.3 Patient Experience Rating:

Interestingly, our analysis showed a positive relationship between patient experience ratings and support for regulation. This counterintuitive finding suggests that even in areas where patients report better experiences, there is still support for increased regulation.



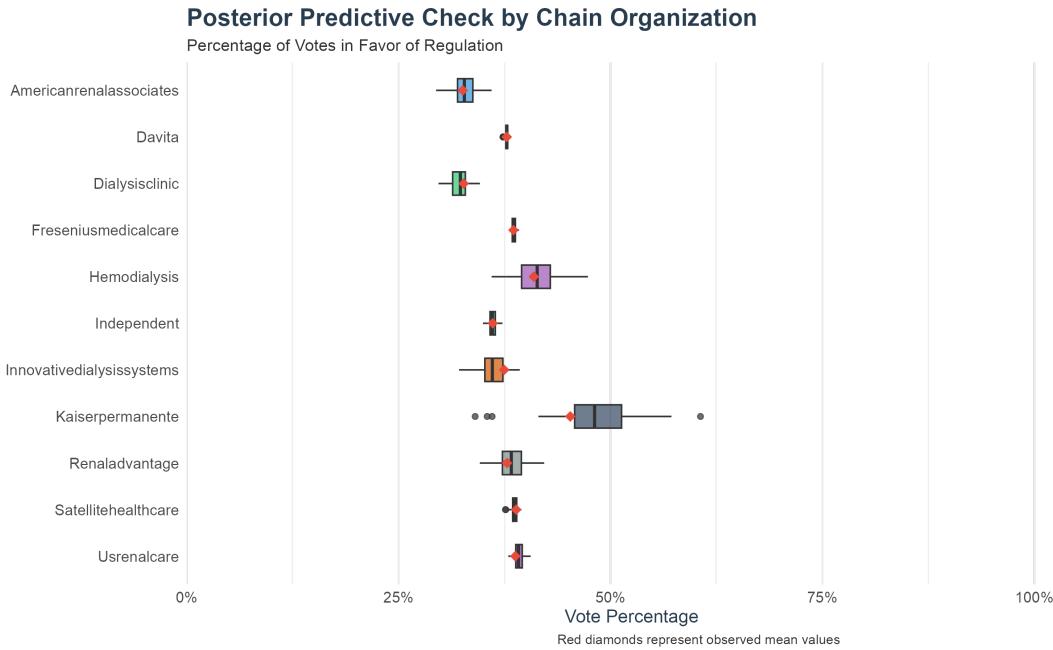
Effect of Patient Experience Rating on Predicted Vote Percentage

5.2.4 Five Star Rating and Stations per Facility:

The associations of these metrics on voting behavior were weaker compared to staff ratings, mortality rates, and patient experience ratings. The estimated effects suggested by the posterior predictive checks were clustered around zero.

5.2.5 Chain Organization Effects:

The posterior predictive check for chain organizations showed varying levels of support for regulation across different dialysis chains, indicating that organizational factors may play a role in shaping public opinion or voting behavior.



Posterior Predictive Check for Chain Organizations

5.2.6 Facility Size Considerations:

The number of dialysis stations (a proxy for facility size) showed a slight positive relationship with voting in favor of regulation, suggesting that areas with larger facilities might be more supportive of increased oversight.

5.3 Challenges and Limitations

Our analysis faced several key constraints that warrant consideration:

1. **Data Granularity:** Facility-level quality metrics were paired with city/county-level voting data, potentially obscuring finer-grained relationships and risking ecological fallacy.
2. **Temporal Dynamics:** The model assumes immediate effects of facility metrics on voting behavior, potentially overlooking lagged effects or longer-term trends.
3. **Confounding Factors:** Unmeasured variables such as socioeconomic factors, political leanings, or media coverage may influence both facility quality and voting patterns.
4. **Causal Interpretation:** While our model reveals associations, causal relationships cannot be inferred without further analysis.
5. **Measurement and Data Quality:** Quality metrics and aggregated voting data may not perfectly capture true care quality or individual voting behavior, introducing potential measurement error.
6. **External Validity:** Findings from California may not generalize to regions with different political, demographic, or healthcare landscapes.

These limitations highlight opportunities for future research.

6 Future Work and Next Steps

Our Bayesian analysis of dialysis facility quality metrics and voting patterns on dialysis-related propositions in California has provided valuable insights. However, several avenues for future research and methodological improvements remain:

1. **Granular Data Collection:**
 - Incorporate sub-city (such as census tract) voting data to mitigate ecological fallacy risks.
 - Collect more detailed patient-level data to better understand the link between personal experiences and voting behavior.
2. **Temporal Analysis:**
 - Investigate how changes in facility quality over time correlate with shifts in voting behavior.
3. **Additional Variables:**
 - Incorporate socioeconomic factors, political leanings, and media coverage data as potential confounders.
4. **Causal Inference:**
 - Explore natural experiments, such as sudden changes in facility ownership or closures, and their impact on voting behavior.

7 Statement of Contribution

7.1 Collaboration

As a group, each member proposed a project proposal including sources, topics, and possible analyses. We met frequently during the weeks of the project to align on current progress and next steps.

7.2 Iris Lin

Managed project logistics, including status updates and coordination of group meetings. Performed research into ballot measurement data, census data, and demographics data. Prepared the final report and ensured project rubric was followed - including code review and extensive documentation.

7.3 Kasra Afzali

Performed exploratory data analysis and created visualizations to understand the data and the relationships between variables, with a focus on geographic data and representation. Built a database of the facilities as a Python class for future use and analysis.

7.4 Michael Light

Led the data processing and analysis efforts, including developing automated scripts for importing and cleaning CMS dialysis facility data. Constructed and refined multiple Bayesian multilevel models using R, conducting in-depth analysis of facility metrics and voting patterns. Managed the technical aspects of report creation, synthesizing findings into a coherent narrative and creating visualizations to effectively communicate the project's outcomes.