Comprehensive Evaluation of Surgical Site Infection Trends Across California Healthcare Facilities

Kiran Devihosur   Sayan Roy   Rahul Shetty   Venkata Sairam Yanamandra

Sunday, 17 December 2023

1 Abstract

The occurrence, prediction, and management of surgical site infections (SSIs) in hospitals across California are the primary focus of this study. Our analysis uses statistical methods to test five hypotheses related to hospital size, types of surgical procedures, and infection control success using the comprehensive dataset from the California Department of Public Health. The dataset includes data points like hospital characteristics, infection rates, and procedure counts. Complex observations, such as the relationship between hospital bed capacity and average SIR, differences in SIR among surgical procedures, and hospitals' ability to reach the 2020 SSI reduction targets, are revealed by our data. These findings have important ramifications for healthcare policy and hospital management methods, in addition to adding to our knowledge of SSIs in various hospital settings.

2 Introduction

Surgical Site Infections (SSIs) continue to affect patient outcomes and healthcare costs. SSIs affect patients and healthcare systems, as this paper will demonstrate. This research examines current SSI knowledge to highlight the need to solve this issue. Background Surgical site infections (SSIs) are common after surgery, causing harm to patients and healthcare providers. These infections might be superficial or deep-seated. SSIs cause pain, lengthier hospital stays, and higher morbidity and mortality. Healthcare expenditures for SSIs are rising, making them costly. Infections in the surgical site are common after surgery. These infections range from superficial skin infections to more serious subcutaneous, organ, or implanted problems. Despite sterile methods and antibiotic prophylaxis, surgical site infections (SSIs) remain, highlighting the need for hospital surveillance and improvement.

SSIs in California hospitals are the focus of this research paper. The study also seeks to uncover infection-causing factors. This investigation is important because surgical site infections (SSIs) can lead to longer hospital stays, higher readmission rates, and even death. In a world of rising healthcare costs and a focus on patient safety and quality, understanding surgical site infections (SSIs) is crucial.

In hospital infection control systems, SSI monitoring and mitigation have been crucial. Recent advances in data collecting and analysis have provided new paths for understanding and managing illnesses. These advances could help us understand infections and develop better prevention methods. The literature has mostly examined specific surgical techniques or patient demographics.

The present study endeavors to make a valuable contribution to the wider discourse surrounding healthcare quality and patient safety. This study aims to conduct an analysis of data obtained from general acute care hospitals in California in order to identify patterns and risk factors associated with surgical site infections (SSIs). The findings of this research endeavor have the potential to provide valuable insights that can be utilized to inform both policy-making decisions and clinical practices in the field of healthcare. The exploration of this research endeavor offers valuable insights into the potential avenues for enhancing surgical procedures, improving patient outcomes, and optimizing resource allocation within healthcare settings.

3 Data

This study examines 2022 California acute care hospital Surgical Site Infections (SSIs) using Data.gov's "Surgical Site Infections Data". This dataset contains complete SSI data for analysis. The dataset under evaluation is essential for understanding surgical site infections (SSIs) across a variety of surgical procedures.

The process of data collection is a fundamental aspect of this section.Information and evidence are collected to support research objectives and answer questions. The National Healthcare Safety Network helped California general acute care hospitals collect and submit data. CDC securely manages the National Healthcare Safety Network (NHSN), an internet-based surveillance system. Hospitals meticulously collect and publish data on healthcare-associated infections, focusing on surgical site infections. These reports use common criteria and processes to collect data consistently and accurately. Healthcare facilities must standardize data gathering to ensure consistency and enable meaningful comparisons. Healthcare facilities can improve data comparability and analysis by standardizing data gathering. Standardization is necessary for evidence-based decision-making and healthcare quality improvement.

Healthcare data gathering is extensive and dependable via the National Healthcare Safety Network (NHSN). However, biases may affect the accuracy and validity of this system's data. These biases should be considered while reading NHSN data.

1. Reporting Bias: Hospitals may report Surgical Site Infections more or less carefully and accurately. Staffing and infection control resources contribute to these differences.
2. Selection Bias: This dataset may be biased because it only includes data from acute care hospitals, excluding other healthcare facilities that perform surgical procedures. conducted.
3. Detection Bias: Variations in Surveillance Intensity and Diagnostic Practices on Surgical Site Infections Correctly identifying surgical site infections requires intensive surveillance and diagnostics. However, hospital-specific deficiencies in these criteria can cause detection bias, reducing SSI data reliability and comparability. This paper investigates how monitoring intensity and diagnostic techniques affect detection bias in SSI identification. SSI identification errors owing to surveillance intensity and diagnostic techniques are called detection bias. Hospitals' diligence in monitoring SSIs is called surveillance intensity.

During the data cleaning process, we identified specific columns in the dataset that were deemed unnecessary for our research and consequently removed. It is essential to optimize the dataset by prioritizing the most pertinent information. The columns that have been eliminated are:

1. HAI: The content of this column, which pertained to healthcare-associated infections, was considered irrelevant to the precise objective of our study.
2. State: The inclusion of the state column was unnecessary in our investigation as we solely examined hospitals in California.
3. Year: The year column has been excluded as our research is limited to data from a certain year, rendering this column insignificant.
4. Notes: Supplementary remarks or unstructured data that may not directly add to the quantitative analysis were also excluded.

In conducting a comprehensive analysis, it is crucial to identify and consider the key features that are relevant to the research topic. These features serve as the foundation for the analysis and provide valuable insights into the subject matter. This paper aims to outline the important features that should be The dataset encompasses a multitude of significant features, or columns, that are of utmost importance for the subsequent analysis.

The comprehension of these characteristics is imperative in order to conduct a precise analysis of the data and derive significant conclusions. The objective of this analysis is to identify recurring patterns, evaluate potential risk factors, and assess the performance of hospitals in preventing surgical site infections (SSIs). By doing so, this study aims to make a valuable contribution to ongoing endeavors aimed at improving patient safety and the overall quality of healthcare.

Variable	Description
County	Indicates the specific county in which the hospital is situated, providing geographical context.
Operative Procedure	Provides detailed information about a surgical procedure that is essential for comprehending different surgical risk profiles.
Facility ID	A distinct identifier assigned to each hospital, allowing for comprehensive analysis at the hospital level.
Facility Name	Identifies individual hospitals, enabling focused case studies and comparisons.
Hospital Category for Risk Adjustment	Used to classify hospitals based on their risk adjustment in analysis. Includes categories such as 'Acute Care Hospital' and 'Critical Access Hospital'.
Facility Type	Categorizes hospitals based on features such as 'Major Teaching', 'Pediatric', or 'Community Hospital', providing valuable information about the hospitals.
Procedure Count	Refers to the total number of individual operations performed, crucial for assessing the extent of surgical activity.
Reported Infections	Represents the quantity of documented Surgical Site Infections (SSIs), providing insight into the frequency of these occurrences.
Infections Predicted	The anticipated quantity of SSIs determined using standardized models, valuable for comparing observed rates to expected rates.
Standardized Infection Ratio	A significant indicator used to evaluate the success of infection control. Compares the actual number of SSIs to the expected number, accounting for risk variables.
SIR 95% Confidence Interval Lower Limit	The lower bound of the SIR's 95% confidence interval, represents the minimal estimated value, considering variability and precision.
SIR 95% Confidence Interval Upper Limit	The upper bound of the SIR's 95% confidence interval, indicates the highest likely result, considering data variability and precision.
Comparison	Provides a detailed examination in relation to other data or benchmarks.
Met 2020 Goal	Indicates if the hospital achieved the 2022 SSI reduction targets based on the goals set in 2020.
SIR 2015	The SIR value computed for the year 2015, facilitating time-based comparisons.

Table 1: Summary of Dataset Variables

4 Methods

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process. It involves the initial exploration and examination of a dataset to gain insights and understand the underlying patterns and relationships within the data. EDA serves as a foundation for further statistical analysis and

The exploratory data analysis (EDA) phase played a pivotal role in comprehending the dataset and priming it for more extensive examination. In the present study, a series of significant modifications and analyses were undertaken.

1. Data Cleaning and Preprocessing: - Standardization of Column Names: In order to ensure uniformity and facilitate ease of access, the column names were subjected to a standardization process. This involved converting all column names to lower case and replacing any spaces with underscores.

    - Elimination of Repetitions: In order to maintain the integrity and distinctiveness of the dataset, any instances of duplicate entries were detected and subsequently eliminated.
    - Treatment of Missing Values: In order to ensure the integrity of the analysis, any rows containing missing values in key columns were identified and subsequently excluded from the dataset.
    - Data Filtering: In order to ensure the relevance of the data used in this study, a filtering process was implemented. Any data points that were deemed irrelevant, such as those pertaining

to unknown counties or non-relevant surgical procedures, were excluded from the analysis. This step was crucial in order to maintain the integrity and accuracy of the findings.

- Data Type Conversion: In order to facilitate quantitative analysis, it was necessary to convert certain columns to their appropriate data types. This involved converting categorical variables into numerical formats.

2. Identification of Numerical Columns: - In order to conduct a quantitative analysis pertaining to the study of Surgical Site Infections (SSIs), specific columns such as 'facility_id', 'procedure_count', and 'infections_reported' were carefully selected for examination. These columns were chosen due to their direct relevance to the research objectives.

3. Visualization: - In order to gain a better understanding of the distribution of key variables, histograms were created for each numerical column. To enhance the visual representation, Kernel Density Estimates (KDE) were incorporated into the histograms. This technique allowed for a more comprehensive visualization of the data. The analysis conducted facilitated comprehension of the dispersion, measures of central tendency, and potential skewness exhibited by the data.

Analytical techniques are a crucial aspect of research and data analysis. These techniques are employed to examine and interpret data in order to derive meaningful insights and draw valid conclusions. In this paper,

1. Descriptive Statistics: - The present study employed descriptive statistics as a means of summarizing the data. Specifically, fundamental statistical measures such as the mean, median, and standard deviation were computed. These measures provide valuable insights into the central tendency and dispersion of the dataset under investigation. By employing these statistical techniques, a comprehensive overview of the data was obtained, facilitating a more nuanced understanding of the underlying patterns and characteristics.

2. Distribution Analysis: - The examination of histograms with Kernel Density Estimation (KDE) has yielded valuable insights into the distribution patterns of variables such as infection rates and procedure counts. The process of identifying any skewness or outliers in the data was deemed essential.

3. Binary Encoding of Categorical Variables: - The conversion of categorical variables, such as 'met_2020_goal', into binary format was performed in order to render them appropriate for statistical analysis.

4. Hypothesis Testing: - The purpose of this study is to conduct hypothesis testing in order to examine specific hypotheses pertaining to the prevalence of surgical site infections (SSIs) and the factors that influence their occurrence. The hypotheses will be formulated based on existing literature and will be tested using appropriate statistical methods. The findings from this analysis will contribute to the understanding of SSIs and provide valuable insights for healthcare professionals and policymakers in their efforts to prevent and manage these infections. In this study, we aim to conduct a comparative analysis of SSI (Surgical Site Infection) rates among various categories of hospitals. The objective is to examine the potential variations in SSI rates across different types of healthcare facilities. By exploring this topic, we hope to contribute to the existing body of knowledge on hospital-associated infections and provide insights into the factors that may influence SSI rates in diverse hospital settings.

5. Correlation Analysis: - The present study undertook an investigation into the relationships between multiple numerical variables with the aim of identifying potential associations. This analysis involved examining the extent to which these variables were correlated with one another, thereby providing insights into the potential connections that may exist between them. By employing correlation analysis, the study sought to uncover any statistically significant relationships that could shed light on the interdependencies and associations among the variables under investigation.

4.1 Data Analysis

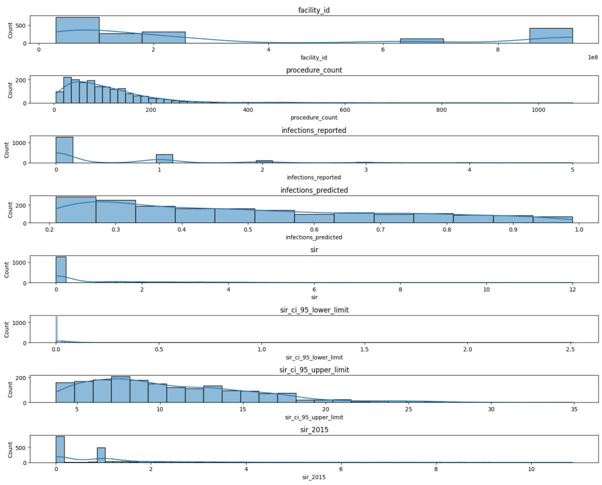

Figure 1: Distribution of Numerical Variables Related to Surgical Site Infections in California Acute Care Hospitals.

Histograms depict California acute care hospital SSI numerical factors. Data input frequency for unique-identifier facilities is shown in the histogram. Positively skewed distribution indicates few institutions have high ID numbers and many have low IDs. Newer facilities may have higher IDs. Right-skewed operation frequency distributions limit procedures for most entries. Many institutions undertake few procedures, according to statistics.Facility-specific SSI histogram. Most facilities reported no infections and a few more, skewing right. Infection modeling: Right-skewed. Models predict few infections for most low-value surgeries.Favorable conditions lead to low Standardized Infection Ratios in most institutions. SSIs may exceed expectations.The 95% confidence interval's right-skewed lower limit suggests that most facilities have a low SIR limit and few have a higher limit.The highest 95% Standardised Incidence Ratio confidence interval is shown below. Although most facilities have a lower top limit, the distribution is less skewed. The distribution's long tail reflects more restrictive facilities. The highest 95% Standardised Incidence Ratio confidence interval is shown below. Although most facilities have a lower top limit, the distribution is less skewed. The distribution's long tail reflects more restrictive facilities.Standardised Infection Ratio 2015 SIRs vary by institution, but most are near zero. Kernel density-estimated histograms show most institutions have fewer operations, infections, and SIRs. The skewness of these distributions means a few institutions manage most operation counts, infection reports, and SIR values.

Transformed surgical site infection variable histograms are shown. Several statistical analyses use square root and logarithm transformations to regulate skewness and normality. interpreting each plot Procedure count SRT dispersion histogram. SRT reduces positive data skewness. Transformers lower high values, but most data is still among lower procedure counts, making it more normalized but with correct skewness.This histogram demonstrates published infection square root transformation. As most institutions have few diseases, the distribution is right-skewed and peaks around zero. No transformation normalized it.Infection_predicted_log histogram: The estimated infections were logarithmically modified by data skewness. Highs replace lows in log transformation. It is more equally distributed than the data despite its right skew.Standardized Infection Ratio logarithmically. Delete outliers and positive skewness from SIR data. Peaks at 0 imply high frequencies of low SIR values with tails rising.SI_95_lower_limit_log histogram: Logarithmically converted SIR 95% confidence interval bottom limit histogram. SIR confidence interval lower limits are low for most facilities due to the right-skewed distribution with a strong peak near 0.SI_95_upper_limit_log histogram: Normalizes distribution using log-transformed SIR 95% confidence interval (upper limit). Logification symmetricalized this distribution by reducing skewness and stabilizing variance.Normal data is assumed for linear regression and histograms. More data processing or non-parametric statistical methods without normality may be needed to correct transformation-induced skewness.

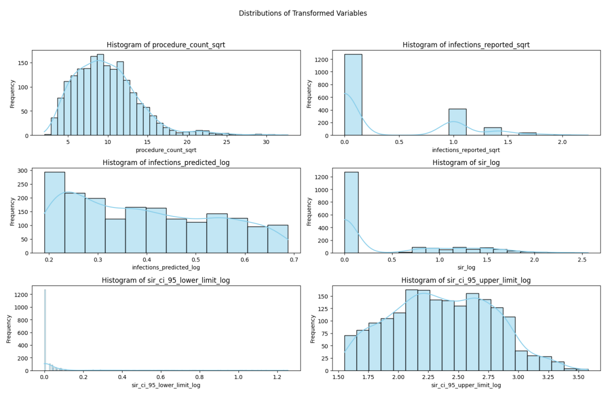

Figure 2: Distributions of Transformed Variables Related to Surgical Procedures and Infections

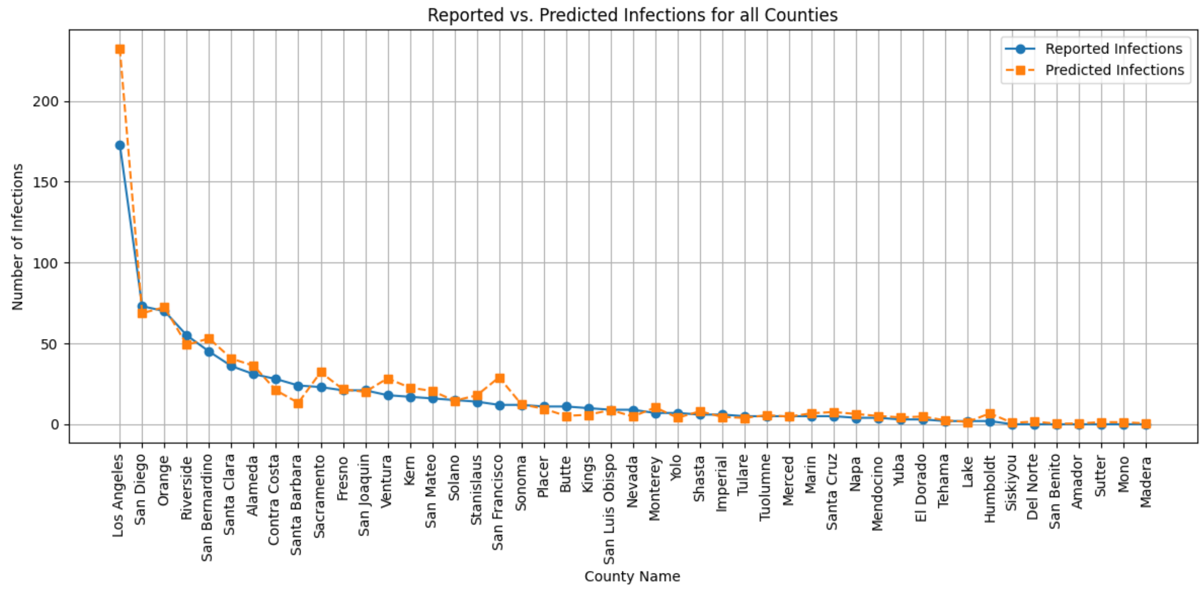

Figure 3: Comparison of Reported vs. Predicted Surgical Site Infections across California Counties.

This chart compares California county SSIs reported and projected. Actual infections are blue, while expected infections for each county are orange. See the plot below.Los Angeles County has the most reported and predicted infections, indicating a higher SSI rate due to population and surgery.Differences between reported and expected values Los Angeles, San Diego, and other counties have unexpected SSIs. This disparity may suggest surgical or postoperative complications.Most counties have close infection records and predictions, indicating predictive effectiveness. The reporting system and model's SSI estimation accuracy are proven by this alignment.Distribution of Data: Most counties predicted fewer than 50 SSIs after the initial few. This may be due to hospital resource allocation, population density, and state-wide healthcare access.County-specific analysis: In counties with high infection rates, county-specific SSI data visualization can help allocate resources, target treatments, and improve surgical care and infection control.Overall trends: Infections are decreasing from larger to smaller counties, suggesting that populated places with higher healthcare resources have more SSIs. Patient volume, hospital capacity, and procedure complexity may cause this.

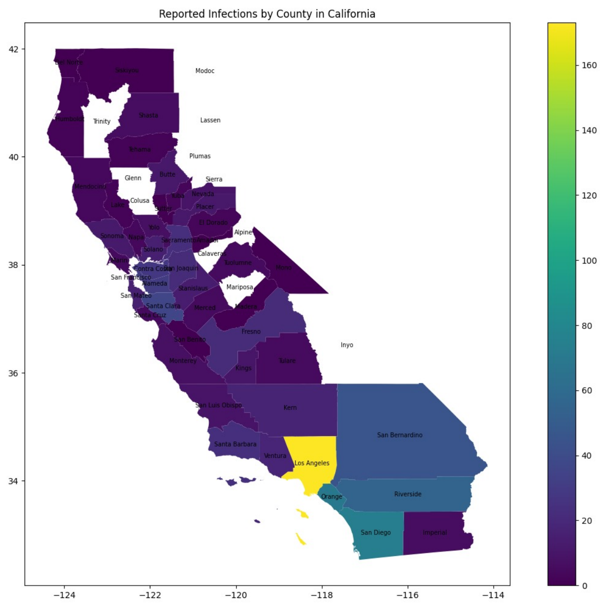

Figure 4: Choropleth Map of Reported Surgical Site Infections by County in California.

The choropleth map shows California county-level Surgical Site Infections (SSIs). Darker tones on the map imply more infections, whereas lighter shades indicate less. As California's most populous county, Los Angeles County has the most SSIs, indicated in yellow. Lighter blue counties like Orange and San Diego have less illnesses. This pattern implies that SSIs are concentrated in densely populated areas due to higher hospital activity and patient flow.Most counties, especially northern ones, have low SSI rates, as shown by the dark purple to black coloration. These regions may have lower population density, less healthcare facilities, or better infection control.

Geographic trends and public health intervention areas are clearly shown on the map. It helps public health professionals and lawmakers allocate resources and develop SSI prevention initiatives across the state.

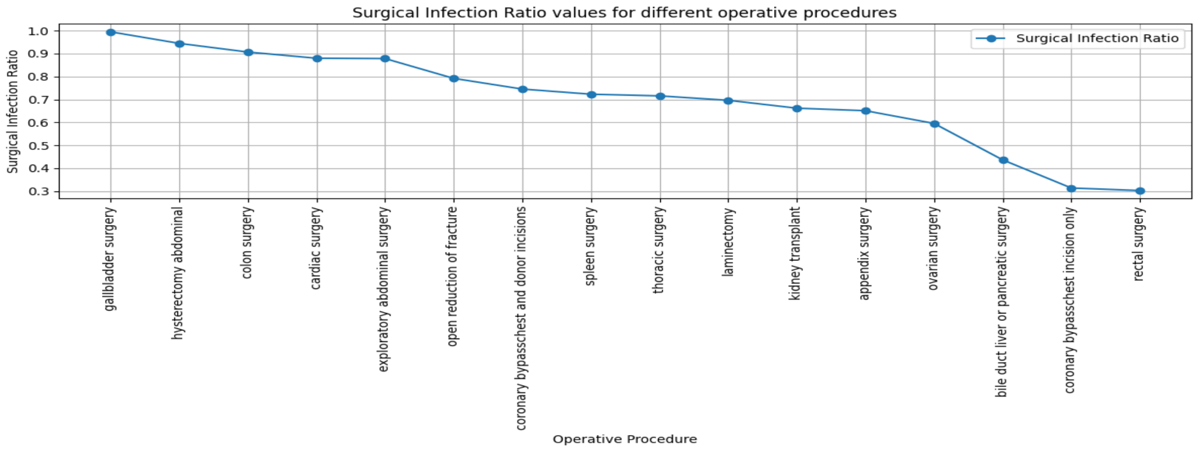

Figure 5: Surgical Infection Ratio Across California Counties by Operative Procedure

The graphic displays the surgical infection ratio (SIR) data for several operating procedures in different counties of California. SIR examines the difference between expected and actual surgical site infections (SSIs), while taking risk factors into account. As we move from the left side of the x-axis to the right, we can see that the SIR values of the counties listed later on are lower. The highest SIR rating, near 1, is for Alameda County, which means that its SSI rate is almost the same as the national average or baseline for anticipated SSIs. The counties of Kern, Mendocino, and El Dorado have higher SIRs, suggesting an unexpected number of SSIs, as we move to the right. The median SIR level in the dataset is along the county lines of Napa, Lake, and Los Angeles. San Mateo, Tehama, and Ventura counties all had lower-than-anticipated SSI values, and the trend has persisted. The counties with the lowest SIRs, which means fewer SSIs than anticipated, were Marin, Yuba, Monterey, San Francisco, and Humboldt. This might be because of local factors that are influencing results, such as reduced surgery volumes, better infection management, or both.

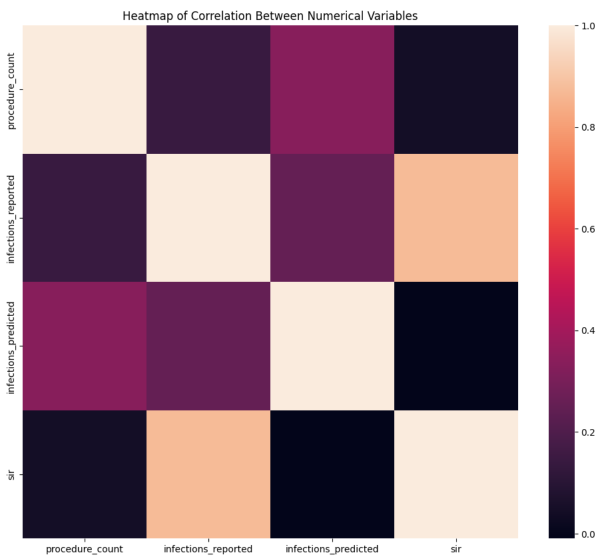

Figure 6: Heatmap of Correlation Between Numerical Variables

The heatmap demonstrates the association between surgical procedures and infections. Each square in the heatmap shows the strength and direction of the association between two variables, with the correlation coefficient values on the right as color intensity and scale. Darker purple indicates a higher negative correlation, beige to pink a positive correlation, and black no association.Starting from the top left corner, 'procedure_count' and 'infections_reported' have a strong positive association, which makes sense since more procedures may lead to more infections. The positive connection between 'procedure_count' and 'infections_predicted' suggests that predictive models foresee more illnesses with more operations.The association between 'procedure_count' and 'sir' is weaker, suggesting that the number of procedures does not directly affect the standardized infection ratio. There is a positive correlation between 'infections_reported' and 'infections_predicted', indicating that predictions match reported facts. However, both 'infections_reported' and 'infections_predicted' have negative correlations with 'sir', suggesting that larger infection counts may lower the standardized infection ratio, possibly due to the ratio's correction for procedures.

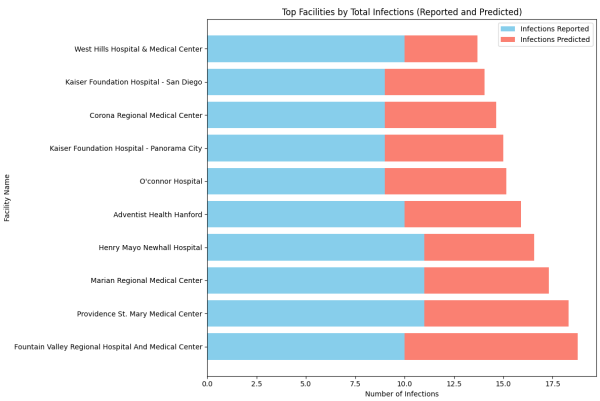

Figure 7: Top Facilities by Total Infections (Reported and Predicted)

The bar chart shows reported and expected surgical site infections (SSIs) at various hospitals. The y-axis lists facilities while the x-axis shows infections. Blue bars show reported infections, whereas red bars show projected infections for each facility. The chart shows Fountain Valley Regional Hospital And Medical Center has the most reported and forecasted infections. This may indicate a larger facility or more surgeries. West Hills Hospital & Medical Center had the lowest counts but a large gap between reported and projected infections, with reported being higher. The figure shows that many facilities have more infections than predicted, which may indicate greater infection rates than predicted by models or benchmarks.This illustration clearly compares reported results to projections, which can help evaluate infection control strategies and identify areas for surgical and postoperative care improvement.

This diagram shows violin graphs for surgical and infectious numerical variables. Violin plots show the distribution of data over multiple levels of a categorical variable and its density at different values.Procedure-count violin plot: A moderate number of procedures is most typical, as shown by this plot's peak. The distribution declines as operation counts grow, suggesting high procedure counts are rare.Violin Plot of infections_reported: Reports of illnesses are clustered at low numbers, with a lengthy tail at higher numbers. Most hospitals report a few infections, but fewer report more.Violin Plot of infections_predicted: Like reported infections, anticipated infection numbers are skewed toward lower values but more evenly distributed, showing the predictive model expects a wider range of infection counts. Sir's Violin Plot: The Standardized Infection Ratio (SIR) plot depicts a multi-modal distribution with numerous data points concentrated in peaks.

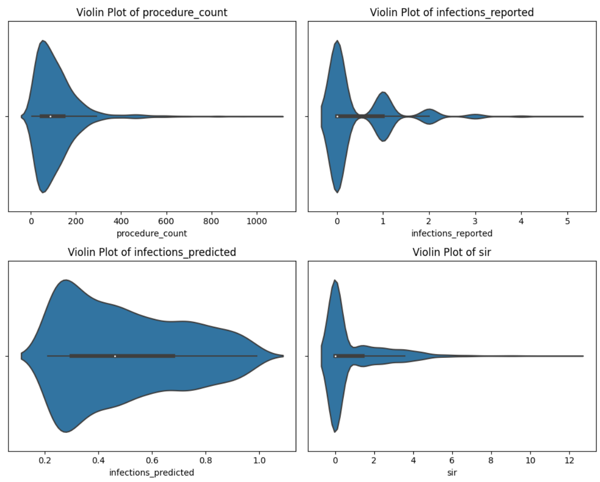

Figure 8: Voilin Plot for Distribution of Procedure Counts and Infection Rates in Healthcare Facilities

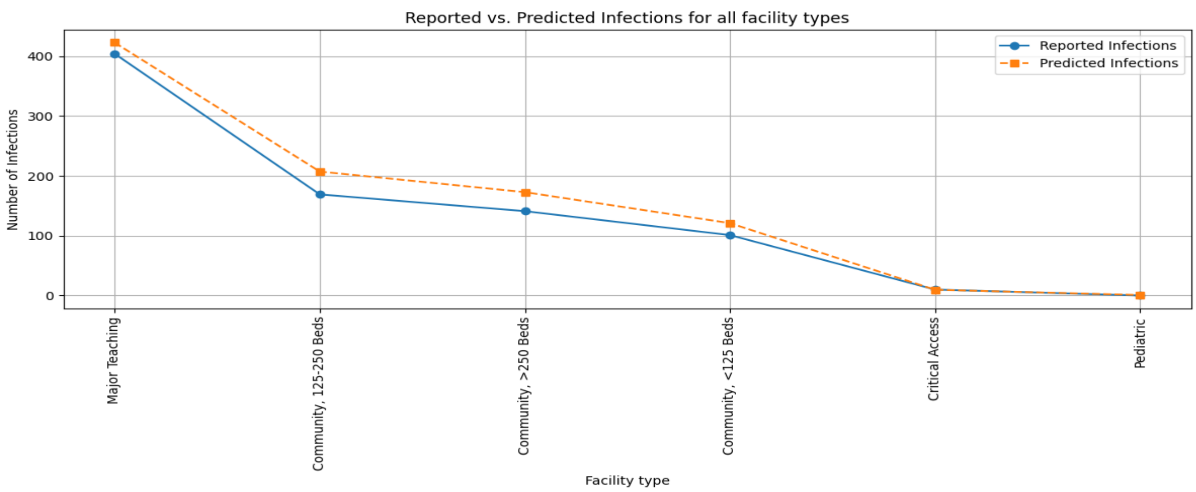

Figure 9: Infection Trends Across Hospital Types: Reported vs. Predicted

The plot contrasts hospital bed numbers with specialization-based reported and expected infections. The blue line shows reported infections, while the orange dashed line shows projected infections at each facility type. Major teaching hospitals, usually large surgical centers, have the most reported and expected infections. High patient volumes and difficult cases are typical of such hospitals. As hospitals with fewer beds (Community greater than 250 Beds, Community 125-250 Beds, and Community less than 125 Beds) execute fewer surgeries, both reported and expected infections decrease. Critical access hospitals, which are smaller and rural, have lower infection rates, as predicted. The lowest rate of infections is in pediatric hospitals, perhaps due to specialized treatment and fewer invasive surgery than general hospitals. The predictive models appear to be well-calibrated to hospital size and type, since reported and anticipated values are close across all facility types. Healthcare administrators and policymakers need this graphic to assess infection burden across hospital categories and plan infection control initiatives.

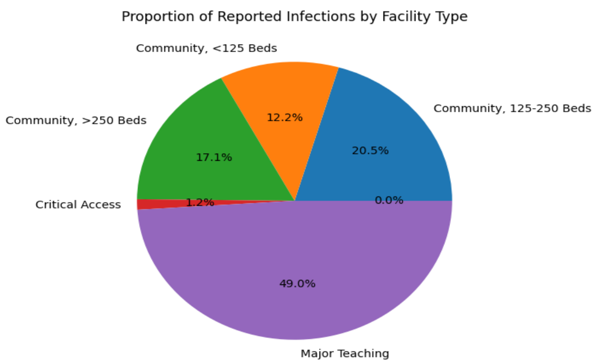

Figure 10: Breakdown of Reported Infections by Hospital Category: A Pie Chart Analysis

The pie chart shows healthcare facility type-specific infection rates. Major teaching hospitals, which account for 49% of infections, are purple. Due to their larger patient volumes and complexity, large academic hospitals may account for roughly half of infections. Community hospitals with 125-250 beds account for 20.5% of infections, followed by those with more than 250 beds at 17.1%. Community hospitals' vast range of services and procedures certainly contributes to their high infection rates. Community hospitals under 125 beds have 12.2% of infections, which may be due to decreased patient throughput and fewer high-risk operations. Critical access hospitals, mainly rural, have a minor share of 1.2%. Pediatric hospitals have a 0.0% infection rate, which may imply a low rate. The figure clearly shows where infections are most common across hospital types, which helps prioritize infection control strategies.

5 Result

This segment conveys the outcomes of a sequence of hypothesis tests pertaining to the Standardized Infection Ratio (SIR), a critical indicator utilized in the evaluation of infection control measures in hospitals. By employing statistical analyses such as t-tests and proportion tests, we examine the relationship between SIR values and variables including hospital size, procedural differences, and predictive accuracy. The subsequent discussions provide a succinct analysis of each hypothesis, clarifying the potential consequences for policies and practices in the healthcare industry.

5.1 Hypothesis 1: Hospitals with a bed capacity above 250 have a lower average SIR in comparison to smaller hospitals

Null Hypothesis H0: Hospitals with a bed capacity above 250 have a lower average SIR in comparison to smaller hospitals.

Alternative Hypothesis H1: Hospitals with a bed capacity above 250 have a higher average Standardized Infection Ratio (SIR) in comparison to smaller hospitals.

The purpose of our investigation was to assess how the size of the hospital influences, as measured by the number of beds, on the Standardized Infection Ratio (SIR). The SIR is a vital metric for assessing rates of surgical site infections. This study aligns with the overarching objective of understanding the factors that impact hospital-acquired infections, thereby offering guidance for improving patient care and hospital management practices.

Approach

By drawing the analogy between the mean SIR (Standardized Infection Ratio) of large hospitals (more than 250 beds) with the smaller hospitals (less than 125 or within 125 to 250 beds) - a two sample t-test was used to conclude the hypothesis. The t-test is statistical test (analysis method) helps to determine the presence of a significant difference between the means of two groups (based on hospital size here).

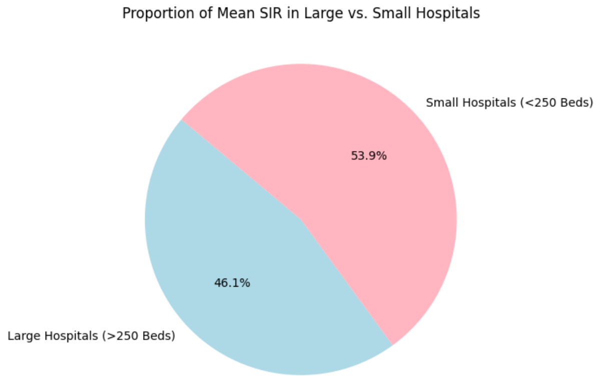

Figure 11: Proportion of mean SIR among large and small hospitals

Findings

- Mean SIR for Large Hospitals (greater than 250 Beds): 0.742
- Mean SIR for Small Hospitals (less than 125 Beds or 125-250 Beds): 0.867
- t-statistic: -1.301
- p-value: 0.194

The t-test yields a P-value of 0.194, which is more than the traditional alpha requirement of 0.05. Given these findings, the difference in the average Standardized Infection Ratio (SIR) between large and small hospitals is unlikely to be statistically significant. As a result, it is not plausible to claim that larger hospitals have a lower average SIR when compared to smaller hospitals. In practice, the findings of this study imply that the number of beds in a hospital has no significant effect on the Standardized Infection Ratio, as demonstrated in the studied data.

5.2 Hypothesis 2: There is a significant difference in the average Surgical Infection Rate (SIR) between colon operations and cesarean sections

Null Hypothesis H0: There is no difference in the Standardized Infection Ratios between Cesarean sections and Colon surgeries.
Alternative Hypothesis H1: There is a difference in the Standardized Infection Ratios between Cesarean sections and Colon surgeries.

In order to enhance our understanding of the factors influencing rates of surgical site infections, we conducted a comparative analysis of different surgical techniques in our study. The primary objective of our investigation was to assess the presence of a statistically significant discrepancy in the Standardized Infection Ratios (SIRs) between Cesarean sections and colon procedures.

Approach

To evaluate this hypothesis, a two-sample t-test was employed. The statistical procedure described herein is designed specifically for the purpose of comparing the means of two distinct groups that are independent of each other. In the present context, a comparison is made between two groups: patients who have undergone Cesarean sections and patients who have undergone colon operations.

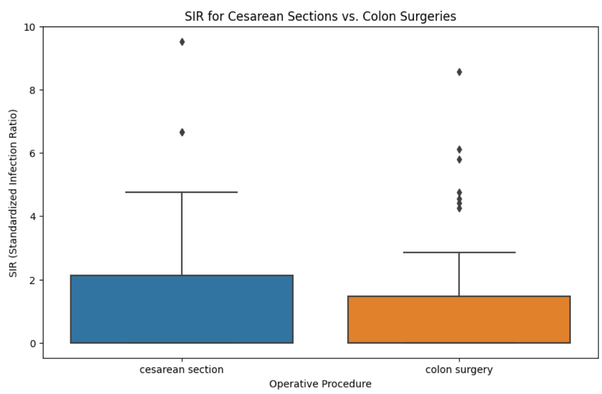

Figure 12: SIR between Cesarean Sections and Colon Surgeries

Findings

- The investigation yielded a p-value of 0.604.

- The t-static value of 0.5199

There is no statistically significant evidence to indicate the presence of a discrepancy in the average SIRs between Cesarean sections and colon operations, as the p-value 0.604 is substantially larger than the conventional alpha criterion of 0.05. To put it plainly, the data does not lend credence to the idea that the SIRs for these two procedures are significantly different.

5.3 Hypothesis 3: A substantial percentage of hospitals have achieved or surpassed the 2020 target for SIR

Null Hypothesis H0: Half of the hospitals achieved the 2020 target for reducing the Standardized Infection Ratio.
Alternative Hypothesis H1: More than Half of the hospitals achieved the 2020 target for reducing the Standardized Infection Ratio.

As part of our analysis to evaluate hospital performance in infection control, we focused on determining the number of institutions that achieved the ambitious 2020 target for the Standardized Infection Ratio (SIR). This objective is a crucial milestone in the healthcare sector, indicating the efficiency of a hospital's infection control protocols.

Approach We utilized statistical analysis to ascertain the fraction of hospitals that successfully attained the 2020 SIR objective. More precisely, we employed a test of proportions to compare the actual proportion of hospitals that achieved the objective with a hypothetical proportion of 50%.

Findings

- Proportion of Hospitals Meeting the 2020 Goal: 68.44%
- The t-statistic is 17.123
- The p-value is 9.93×10^{-66}

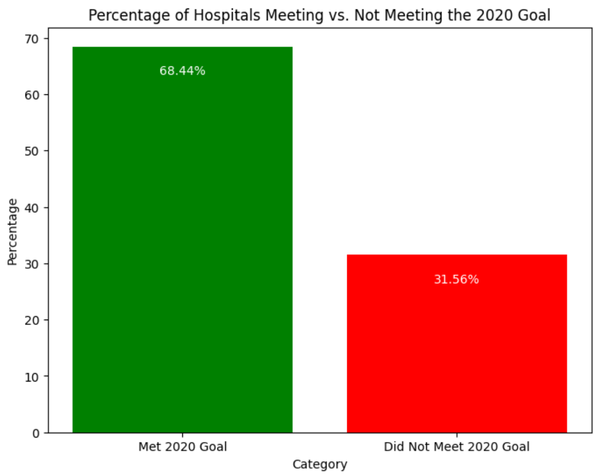

Figure 13: Representation of percentages of Hospital meeting the goals or not

The research resulted in a very significant p-value of roughly 9.93×10^{-66}, which is far less than the conventional alpha threshold of 0.05. The exceptionally small p-value provides strong evidence to reject the null hypothesis that only 50% of hospitals achieved the 2020 SIR objective. The discovery that 68.44% of hospitals have achieved the 2020 SIR objective is evidence of the significant endeavors and progress made in infection control inside these establishments. This suggests that a considerably larger percentage of hospitals have successfully reached the objective in comparison to the hypothetical benchmark of 50%. This result not only confirms the theory but also emphasizes the efficacy of the strategies adopted by these institutions to manage and decrease infection rates.

5.4 Hypothesis 4: A notable disparity exists between the reported cases of Surgical Site Infections (SSIs) and the projected figures derived from known models

Null Hypothesis H0: The mean reported infections are lower than the mean projected infections.
Alternative Hypothesis H1: The mean reported infections are not lower than the mean projected infections.

The objective of our study is to determine if there is a statistically significant disparity between the reported number of surgical site infections (SSIs) and the projected number of SSIs in hospitals. This comparison is essential for comprehending the precision of prediction models and the efficacy of hospital infection control policies.

Approach In order to examine this hypothesis, we used a two-sample t-test. This statistical technique involves comparing the average values of two groups, namely the reported and expected SSIs in the dataset.

Findings

- The t-statistic is -3.409
- The p-value is 0.000666

The negative t-statistic indicates that the mean reported infections are lower than the mean projected infections. The p-value, which is below the normal alpha threshold of 0.05, demonstrates a statistically significant disparity between the reported and expected infection rates. The outcome confirms the hypothesis, demonstrating a significant difference between the reported and projected SSIs. The discrepancy may arise due to several variables, such as the omission of developments in infection control techniques in the predictive models or the possibility of under reporting of illnesses.

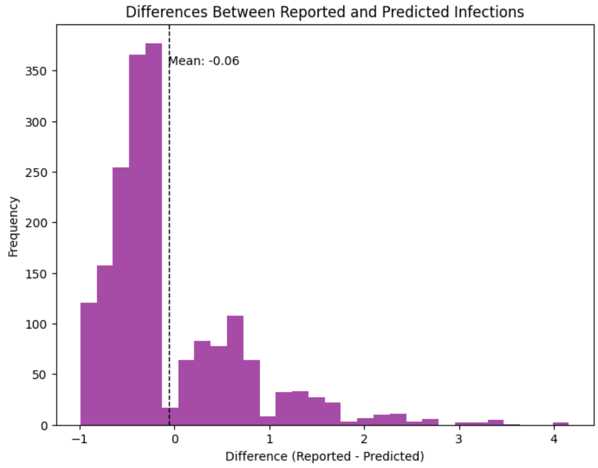

Figure 14: Disparity between Reported and predicted infections

5.5 Statistical Analysis: The distribution of sample means of SIR for stomach procedures follows the Central Limit Theorem

The Central Limit Theorem (CLT) is a crucial statistical theorem that allows us to draw conclusions regarding population parameters. The theory posits that as the sample size expands, the distribution of the sample means will converge towards a normal distribution, irrespective of the form of the population distribution. Understanding this idea is essential for comprehending the behavior of sample means and for carrying out hypothesis testing.

Approach In order to verify the Central Limit Theorem (CLT), we analyzed the characteristics of the distribution of sample means obtained from various sample sizes (10, 30, 50, 100) for the Surgical Infection Rate (SIR) associated with stomach procedures. We generated density plots to visually represent the distribution of sample means and examine the effect of larger sample sizes on the dispersion of the data.

Findings

- The standard deviation of the sample means decreases as the sample size increases, in accordance with the predictions of the Central Limit Theorem (CLT).

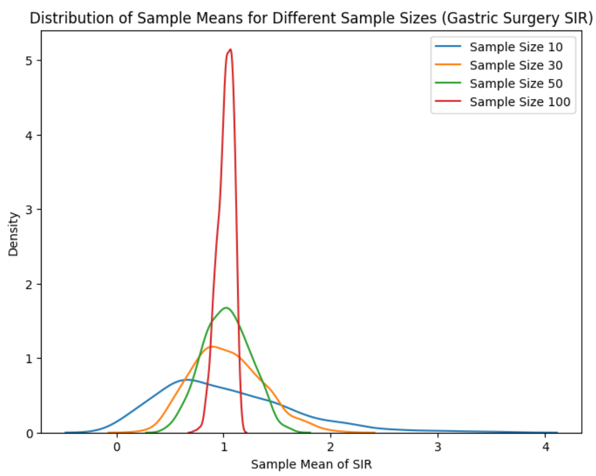

Figure 15: Distributions of sample means across different sample size

- The plots demonstrate that increasing the sample size results in a reduced variability of the sample means. The trend towards a sharper peak suggests a decrease in variability and a greater adherence to a normal distribution.

- The phenomenon of the standard deviation dropping as the sample size increases is a well-known manifestation of the Central Limit Theorem (CLT). The statement confirms the underlying assumption of the theory, which states that as the sample size increases, the distribution of sample means will approach a normal distribution.

The density patterns found across various sample sizes offer empirical evidence for the applicability of the Central Limit Theorem in the context of SIR for stomach procedures. As anticipated by the Central Limit Theorem (CLT), the distribution of sample means tends to approach a normal distribution as the sample size grows larger. This phenomenon indicates less variability and strengthens the theorem's validity for statistical analysis in healthcare data.

6 Conclusion

The study's findings suggest that the size of a hospital and the type of surgical procedure performed do not significantly affect Surgical Site Infection (SSI) rates. Additionally, the data indicates that over half of the hospitals have met or exceeded the SSI reduction targets set for 2020, reflecting positively on the current infection control measures. However, there is a notable trend of hospitals reporting fewer SSIs than predicted, suggesting potential areas for improvement in infection tracking or reporting accuracy. Overall, these results highlight the progress made in infection control while also underscoring the importance of continuous evaluation and refinement of SSI prediction models and reporting practices to further enhance patient safety.

7 References

1. Strategies to Prevent Surgical Site Infections in Acute Care Hospitals: 2014 Update

2. Surgical site infections: epidemiology, microbiology and prevention

3. Surgical site infections post cesarean section and associated risk factors: a retrospective case-control study at a tertiary hospital in Kenya

4. Risk factors for surgical site infection after cesarean delivery: A case-control study

5. Surgical site infection after cesarean section: Implementing 3 changes to improve the quality of patient care

6. Preventing surgical site infection. Where now?

7. Surgical Site Infections