

# TDI HACKATHON

# PROJECT REPORT

## Reducing Surgical Site Infections in California Hospitals

**Team: Data Spartans**

### Executive Summary

Our team developed an interactive Power BI dashboard that analyses Surgical Site Infections (SSIs) across California's healthcare system using comprehensive data from the California Department of Public Health (CDPH) and the National Healthcare Safety Network (NHSN) covering 2021 to 2023. Our analysis reveals critical insights into infection patterns, facility performance, and geographic disparities that inform targeted policy interventions to enhance surgical safety statewide.

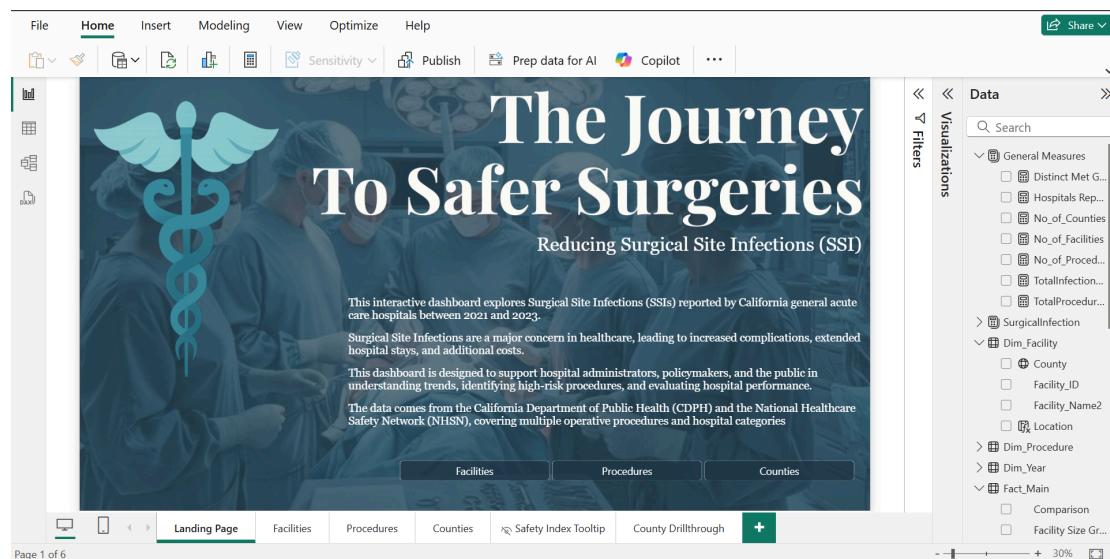


Image: Dashboard Landing Page Screenshot

The dashboard demonstrates that while California has made progress in infection control, significant challenges remain: only **31.81%** of facilities met the 2020 SIR goal, with **24.93%** of facilities classified as unstable in their infection

control performance. The findings identified the need for differentiated approaches based on facility type, procedure risk, and geographic location.

## **Project Objective**

The primary goal was to create a comprehensive data-driven solution that enhances surgical safety across California's healthcare system through advanced infection surveillance and predictive analytics. Our dashboard serves two purposes: analysing historical surgical site infection patterns to identify high-risk facilities, procedures, and geographic areas requiring immediate intervention, and providing actionable insights to guide targeted policy decisions that will reduce infection rates and improve patient outcomes statewide.

## **Data Exploration**

### **Dataset Overview**

The analysis encompasses a comprehensive dataset that merges data from 2021, 2022, and 2023. This consolidated dataset shows:

- **342 healthcare facilities** across California counties
- **11,000 reported infections** spanning multiple procedure types
- **28 operative procedure categories** as mandated by California Health and Safety Code section 1288.55
- **Time period:** 2021-2023, providing three years of comparative data
- **Facility tracking:** Using Electronic Licensing Management System (ELMS) primary keys (Facility ID) maintained by CDPH

## Data Model Architecture

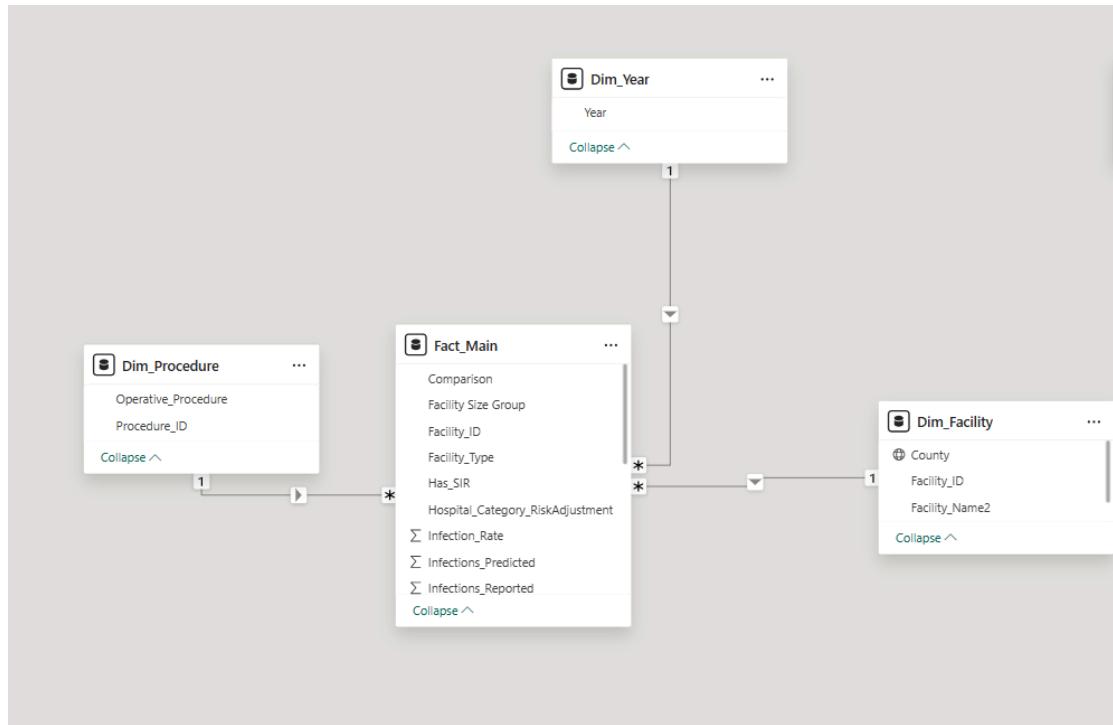


Image: Data Model

A robust star schema data model with three key dimensions and one central fact table was built:

### Dimension Tables:

- **Dim\_Year**: Contains the temporal dimension (2021, 2022, 2023), enabling trend analysis across the three-year study period
- **Dim\_Facility**: Includes facility-level attributes, including Facility\_ID, Facility\_Name, and County locations derived from CDPH licensing data
- **Dim\_Procedure**: Contains all 28 NHSN-defined operative procedure categories with Procedure\_ID and Operative\_Procedure names, from high-volume procedures like Cesarean sections to complex surgeries like heart transplants

**Fact Table (Fact\_Main)**: The central fact table integrates all quantitative measures and categorical attributes:

- **Infection metrics:** Infections\_Reported (actual cases), Infections\_Predicted (risk-adjusted baseline), and Procedure\_Count (surgical volume)
- **Performance indicators:** SIR (Standardized Infection Ratio), SIR\_CI\_95\_Lower\_Limit and SIR\_CI\_95\_Upper\_Limit for confidence intervals
- **Facility characteristics:** Hospital\_Category\_RiskAdjustment (Two NHSN categories), Facility\_Type (six specific hospital types),
- **Quality measures:** Comparison (statistical significance), Met\_2020\_Goal (30% reduction target), Has\_SIR (data completeness indicator)
- **Risk assessment:** Infection rate and SIR\_2015 (historical baseline comparison).

This model enables efficient querying across facility types, procedure categories, temporal periods, and geographic regions while maintaining data integrity through established relationships.

## Key Variables and Metrics

Although the dataset includes 18 columns, below is a summary of the key variables:

1. **Standardized Infection Ratio (SIR):** Primary outcome calculated by dividing observed infections by predicted infections using NHSN's 2015 baseline multivariate logistic regression models
2. **Facility categorisation:** Based on NHSN risk adjustment categories:
  - a. Acute Care Hospital,
  - b. Long-Term Acute Care Hospital (LTAC),
  - c. Critical Access Hospital, and
  - d. Rehabilitation Hospital/Unit
3. **Hospital types:** Major Teaching, Pediatric, Community (by bed size: >250, 125-250, <125), Long-Term Acute Care, Critical Access, and Rehabilitation facilities
4. **Geographic distribution:** County-level analysis using CDPH county codes from facility license applications
5. **Procedure-specific metrics:** Deep incisional and organ/space infections only, excluding superficial SSIs, tracked during hospital admission or readmission

6. **2020 Goal Achievement:** Based on 30% reduction target from 2015 baseline  
(SIR < 0.70)

## Data Limitations

The data exploration revealed several key limitations based on NHSN and CDPH methodologies:

- **Missing SIR data:** 49.25% (8,998) of the records lacked complete SIR reporting due to NHSN criteria requiring at least one predicted infection or CDPH threshold of predicted infections  $\geq 0.2$
- **Confidence interval precision:** SIRs based on smaller numbers of predicted infections have wider confidence intervals, affecting statistical significance determinations
- **Risk adjustment scope:** NHSN risk adjustment models only available for acute care and critical access hospitals
- **Infection definition:** Analysis limited to deep incisional and organ/space infections identified during hospital admission or readmission, excluding superficial SSIs
- **Baseline reference:** All comparisons use 2015 NHSN baseline data, which may not reflect current patient populations or surgical practices

## Preprocessing and Data Handling

### Missing Data Strategy

To ensure the statistical integrity of the NHSN-generated SIR data, all original columns, including SIR estimates and confidence limits, were preserved in their raw form. Rather than altering these expert-calculated fields, we engineered a set of analytic support columns that respect clinical accuracy while unlocking interpretive power.

### Data Standardization

- **Appended the yearly datasets:** 2021, 2022, and 2023 into a single consolidated table for analysis.
- **Facility identification:** We used the Facility\_ID to standardise facilities with differing names but the same ID for consistent facility tracking across reporting periods.

- **Column Removal for Irrelevance:** we removed the **HAI** and **State** columns from the dataset. These columns contained only a single unique value across all records, contributing no variability or analytical value to the study.
- **Removed duplicate:** Duplicates were removed to avoid redundancies across years
- **Trimmed whitespace:** key text fields (e.g., Facility Name) were trimmed to standardize entries
- **Facility Type Standardisation:** To simplify and group similar facility types for easier analysis and visualisation, we standardised the Facility\_Type classifications as follows:
  - **Community, <125 Beds** → *Small Facility*
  - **Community, 125–250 Beds** → *Medium Facility*
  - **Community, >250 Beds** → *Large Facility*
  - **Major Teaching** → *Teaching Hospital*
  - **Pediatric** → *Pediatric Facility*
  - **Critical Access** → *Critical Access*

## **Performance Benchmarking**

Our analytical approach followed NHSN standards:

- **Temporal trends:** Three-year analysis enables identification of sustained improvement or decline patterns
- **SIR performance** was determined using the existing comparison field, which benchmarks each facility's performance as better, worse, or same, relative to the national standard. This ensured consistency with CDC-defined thresholds.
- **Met Goal Classification:** This classification was built on top of the existing Met Goal column:  
Facilities with “Yes” were tagged as Goal Met.  
Facilities with “No” or blanks were grouped as Goal Missed
- **SIR Status** was a simple Yes/No flag used to identify whether a facility had a valid SIR value, ensuring that null or suppressed records were properly handled in analysis.

- **Range Width** was engineered by subtracting the lower 95% CI from the upper 95% CI, producing a measure of statistical spread for each facility's SIR estimate.
- **Stability** was derived from Range Width, with facilities above the 75th percentile classified as “Low Stability” — helping to spotlight hospitals with wider uncertainty in their estimates.
- **Surgical Safety Index (SSI):** To holistically assess how each hospital is performing in terms of surgical infection prevention, we created a Surgical Safety Index. This composite score combines multiple important dimensions: SIR values, performance comparison, and infection burden, to give us a standardised way to identify facilities doing well and those needing urgent attention.

## **Why We Created the Surgical Safety Index**

Surgical Site Infections (SSIs) are influenced by more than just the Standardised Infection Ratio (SIR). While SIR measures how observed infections compare to predicted infections, it doesn't capture other elements like how many infections occurred or how a facility is performing relative to others. So we designed the Surgical Safety Index to:

- Capture multiple risk dimensions in a single, easy-to-read score.
- Help stakeholders quickly compare facility safety levels.
- Support decision-making for interventions and resource allocation.

## **Dashboard Design Methodology**

### **User-Centered Design Approach**

The dashboard was designed with three primary stakeholder groups in mind:

1. **Hospital administrators** requiring facility-specific performance metrics
2. **Policymakers** needing statewide trends and geographic insights
3. **Public health officials** focused on intervention targeting and resource allocation

## **Dashboard Architecture and Visualization**

The dashboard comprises three interconnected pages designed for different analytical perspectives:

**The Facilities Analysis Page:** This page revealed the performance of facilities related to surgical site infections (SSIs) across California. It highlights that out of 342 facilities, over 11,000 infections were reported, with an average Standardised Infection Ratio (SIR) of 0.85. Notably, 24.93% of facilities are categorised as unstable, and the Surgical Safety Index stands at 76.04. This page is designed to help users identify high-risk institutions and recognise trends in infection volume by facility type.

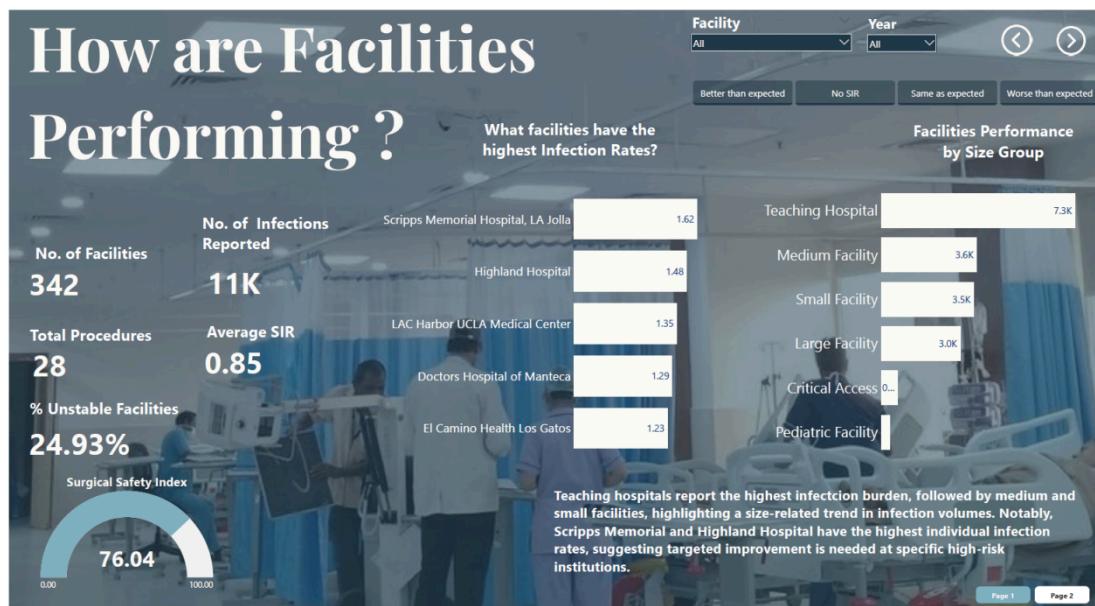
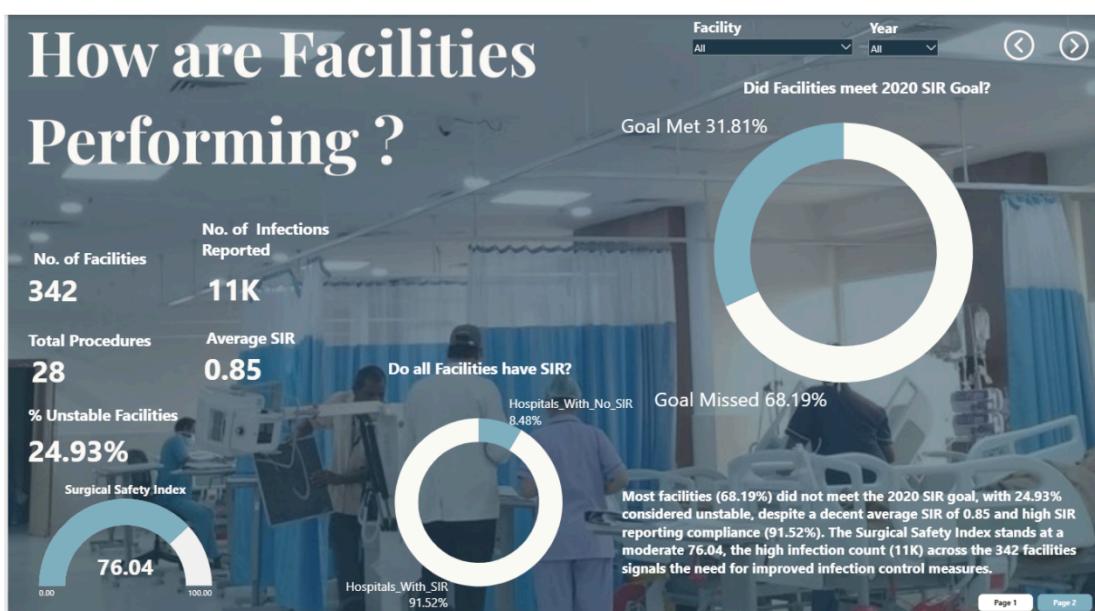
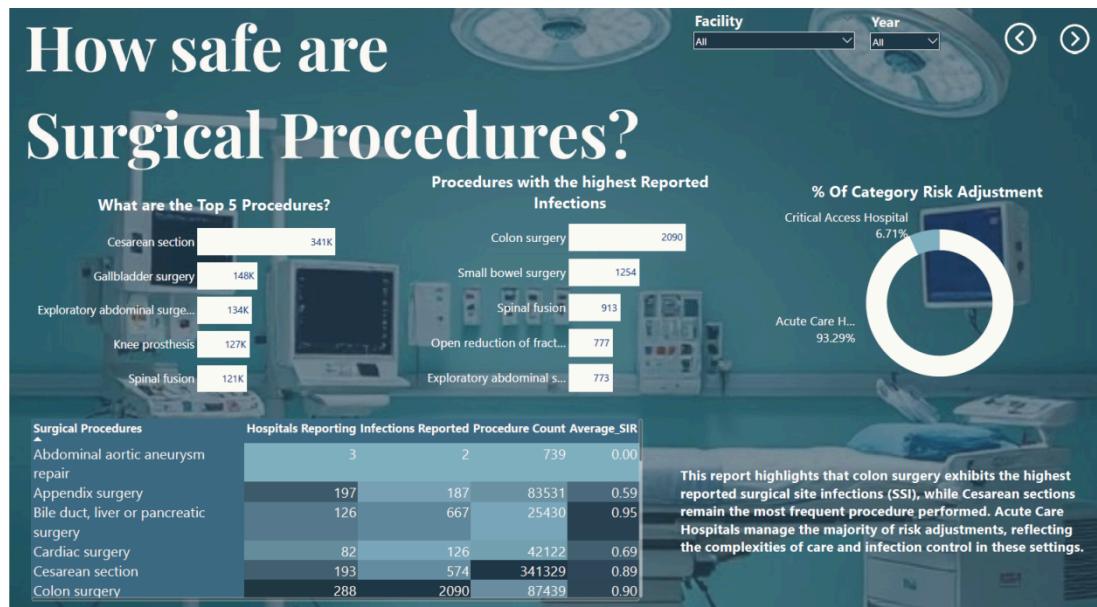


Image: Facilities Analysis Dashboard Screenshot (a)



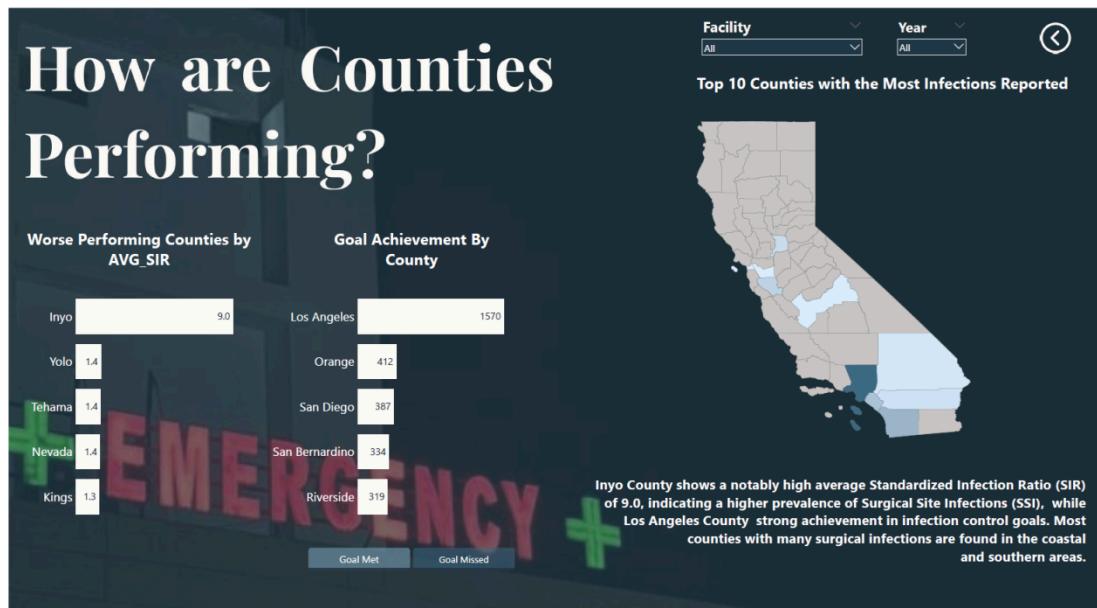
*Image: Facilities Analysis Dashboard Screenshot (b)*

**The Procedures Analysis Page:** This page focuses on evaluating which procedures were safe as well as ones with high infection reporting rates. It was done by examining infection rates and procedure volumes across the hospitals. The analysis highlights colon surgery as the highest-risk procedure and emphasises the significant infection control role of acute care hospitals.



*Image: Procedures Analysis Dashboard Screenshot*

**Counties Analysis Page:** This page evaluates county-level performance in surgical infection control. The page reveals that most infections are concentrated in coastal and southern California, with Inyo County emerging as a key area for intervention and Los Angeles County as a benchmark for strong performance.



*Image: Counties Analysis Dashboard Screenshot*

## Key Insights and Findings

### Overall Performance Landscape

The analysis reveals a healthcare system with significant performance disparities:

- **An average SIR of 0.85** indicates generally effective infection control, but with substantial variation
- **A goal achievement rate of 31.81%** suggests widespread challenges in meeting national standards
- **24.93% unstable facilities** indicates systemic issues requiring targeted intervention
- Our overall average **Surgical Safety Index was 74.06**, which suggests that while many facilities are on track, there's still room for improvement, especially in reducing raw infection counts and improving stability.

### Facility Type Performance Patterns

The analysis uncovered distinct performance patterns using NHSN facility categorisation:

- **Major Teaching hospitals** report the highest infection burden (7.3K infections) but handle complex cases requiring specialized risk adjustment

- **Community hospitals** show variation by bed size: >250 beds (3.6K infections), 125-250 beds, and <125 beds demonstrate different risk profiles
- **Critical Access hospitals** demonstrate better relative performance when adjusted for their rural location and limited resources ( $\leq 25$  beds,  $> 35$  miles from other hospitals)
- **Pediatric hospitals** show distinct infection patterns requiring age-specific intervention approaches

### **Procedure-Specific Risk Profiles**

The dashboard identifies clear procedure-specific risk patterns among the 28 NHSN-defined categories:

- **Colon surgery** leads in reported infections (2,090 cases), indicating need for enhanced protocols in this high-risk category
- **Cesarean sections** remain the highest-volume procedure (341K) with relatively better infection control outcomes
- **Small bowel surgery** and **spinal fusion** show elevated infection rates requiring specialized attention
- **Cardiac surgery** procedures (including coronary bypass) demonstrate varying risk profiles based on complexity
- **Orthopedic procedures** (hip prosthesis, knee prosthesis) show different infection patterns requiring joint-specific protocols

### **Geographic Disparities**

County-level analysis reveals significant geographic variations:

- **Inyo County** shows the highest average SIR (9.0), indicating severe performance challenges
- **Los Angeles County** demonstrates strong goal achievement (1,570 facilities meeting targets)
- **Coastal and southern counties** show concentration of high-infection facilities, suggesting regional factors

### **High-Risk Facility Identification**

The dashboard identifies specific facilities requiring immediate attention:

- **Scripps Memorial Hospital, La Jolla** (SIR: 1.62)
- **Highland Hospital** (SIR: 1.48)
- **LAC Harbor UCLA Medical Center** (SIR: 1.35)

These facilities represent critical intervention opportunities with clear performance gaps.

## **Policy Recommendations**

Based on the findings of the analysis and the CDC [guidelines](#) for preventing SSIs, we propose the following recommendations under short-, medium-, and long-term strategies:

### **Short-Term Actions (0–6 Months)**

#### **1. High-Risk Facility Rapid Response Program**

##### **CDC Alignment: Tailored interventions based on infection rates and facility risk**

Deploy infection control teams to the top 10 high-risk facilities, like Scripps Memorial, Highland Hospital, and LAC Harbour UCLA, to audit CDC practices (timely antibiotics, proper hair removal, normothermia) and provide weekly dashboards and tailored action plans for colon and small bowel surgeries based on infection rates.

#### **2. Academic Medical Center (AMC) Infection Prevention Support**

##### **CDC Alignment: Adjusted expectations based on patient complexity and volume**

Support academic medical centres by implementing risk-adjusted benchmarking for surgical complexity and funding infection control fellowships with SSI reduction training and launching a peer mentorship programme between low- and high-SIR **teaching hospitals**.

## **Medium-Term Strategies (6–18 Months)**

### **3. Procedure-Specific Standardization (CDC Bundle Implementation)**

Implement mandatory SSI prevention bundles for high-risk procedures, including pre-op CHG showers, oral antibiotics with bowel prep, and normothermia tracking for colon surgery; extended post-op monitoring and sterile field maintenance for spinal fusion; and broadened antibiotic protocols with enhanced wound care training for small bowel surgery.

### **4. Regional Disparity Reduction Strategy**

#### **CDC Alignment: Equity-based support and collaborative learning**

Deploy a mobile infection control task force to rural counties like Inyo with high SIR, establish regional collaboratives like the Coastal CA SSI Alliance, and create virtual consultation hubs connecting urban specialists to rural and critical access hospitals.

### **5. Surveillance and Reporting Upgrade Plan**

#### **CDC Alignment: Surveillance integrity is key to prevention**

Mandate NHSN-compliant SIR reporting for all California acute care hospitals and incentivise real-time data systems linked to facility dashboards, while auditing to address 8.48% data gaps.

## **Long-Term Plan (18+ Months)**

### **6. Performance-Linked Payment Reform**

#### **CDC Alignment: Financial accountability tied to infection outcomes**

Implement SSI rate-based reimbursement modifiers using CDC infection ratio thresholds, and incentivise bundle adherence, ensuring rewards go to high-quality, not high-volume care. Also, award top-performing hospitals annually for infection reduction.

## **7. California Statewide SSI Innovation Network (CSSIN)**

### **CDC Alignment: Cross-institutional collaboration and learning**

Launch a quarterly public SSI scorecard showing hospital SIR trends, fund research grants to study California-specific SSI factors, and create an innovation hub sharing effective tools and protocols.

### **Conclusion**

The analysis highlights critical challenges and actionable opportunities in reducing surgical site infections (SSIs) across California. With a statewide SIR of 0.85, 24.93% of facilities were considered unstable, and only 31.81% of facilities currently meet SIR goals. However, identifying high-risk facilities, procedures, and regions enables targeted interventions and room for improvement through systemic and strategic efforts. Policy recommendations focus on immediate interventions to long-term system reforms, supported by a dynamic dashboard for real-time monitoring. Ultimately, this approach offers a path to safer surgeries and better patient outcomes through data-driven decision-making and sustained collaboration.