

# CVS Health Community Access Analysis

## County-Level Health Needs & Clinic Distribution Report

Data Analysis Team

2025-12-09

### Table of contents

<b>Executive Summary</b>	<b>1</b>
Key Findings . . . . .	1
Coverage Gap . . . . .	1
Socioeconomic Inequity . . . . .	1
Health Need Mismatch . . . . .	2
<b>Key Statistics</b>	<b>2</b>
Distribution of Clinic Counts . . . . .	2
Health Burden Distribution . . . . .	3
Clinic Presence vs Health Burden . . . . .	4
SVI Quartiles and Clinic Distribution . . . . .	4
<b>Geographic Expansion Opportunities Map</b>	<b>5</b>
<b>Strategic Recommendations</b>	<b>7</b>
Priority 1: High-Impact Urban Expansion . . . . .	7
Priority 2: Rural High-Need Markets . . . . .	7
Priority 3: Vulnerable Community Access . . . . .	7
Priority 4: Cluster-Based Expansion . . . . .	7
<b>Next Steps</b>	<b>7</b>

## Executive Summary

### Key Findings

#### Coverage Gap

- `r round((df['clinic_count'] == 0).sum() / len(df) * 100, 1)%` of U.S. counties have zero CVS MinuteClinic locations
- Only `r (df['clinic_count'] > 0).sum()` counties (`r round((df['clinic_count'] > 0).sum() / len(df) * 100, 1)%`) have at least one clinic
- Total clinics across all counties: `r int(df['clinic_count'].sum())`

#### Socioeconomic Inequity

- Counties without clinics have higher SVI scores: `r round(df[df['clinic_count']==0]['svi_overall'].mean(), 3)` vs `r round(df[df['clinic_count']>0]['svi_overall'].mean(), 3)`

- Socioeconomic vulnerability: `r round(df[df['clinic_count']==0]['svi_socioeconomic'].mean(), 3)` (no clinics) vs `r round(df[df['clinic_count']>0]['svi_socioeconomic'].mean(), 3)` (with clinics)

### Health Need Mismatch

- Counties without clinics have higher health burden: `r round(df[df['clinic_count']==0]['health_burden_score'].mean(), 2)` vs `r round(df[df['clinic_count']>0]['health_burden_score'].mean(), 2)`
- This indicates sicker populations have less access to CVS services

## Key Statistics

### Distribution of Clinic Counts

```
plt.figure(figsize=(10, 6))
plt.hist(df['clinic_count'], bins=50, edgecolor='black', alpha=0.7, color='steelblue')
plt.axvline(df['clinic_count'].median(), color='red', linestyle='--', linewidth=2,
            label=f'Median: {df["clinic_count"].median():.1f}')
plt.xlabel('Number of Clinics')
plt.ylabel('Number of Counties')
plt.title('Distribution of Clinic Counts', fontweight='bold')
plt.legend()
plt.grid(alpha=0.3)
plt.show()
```

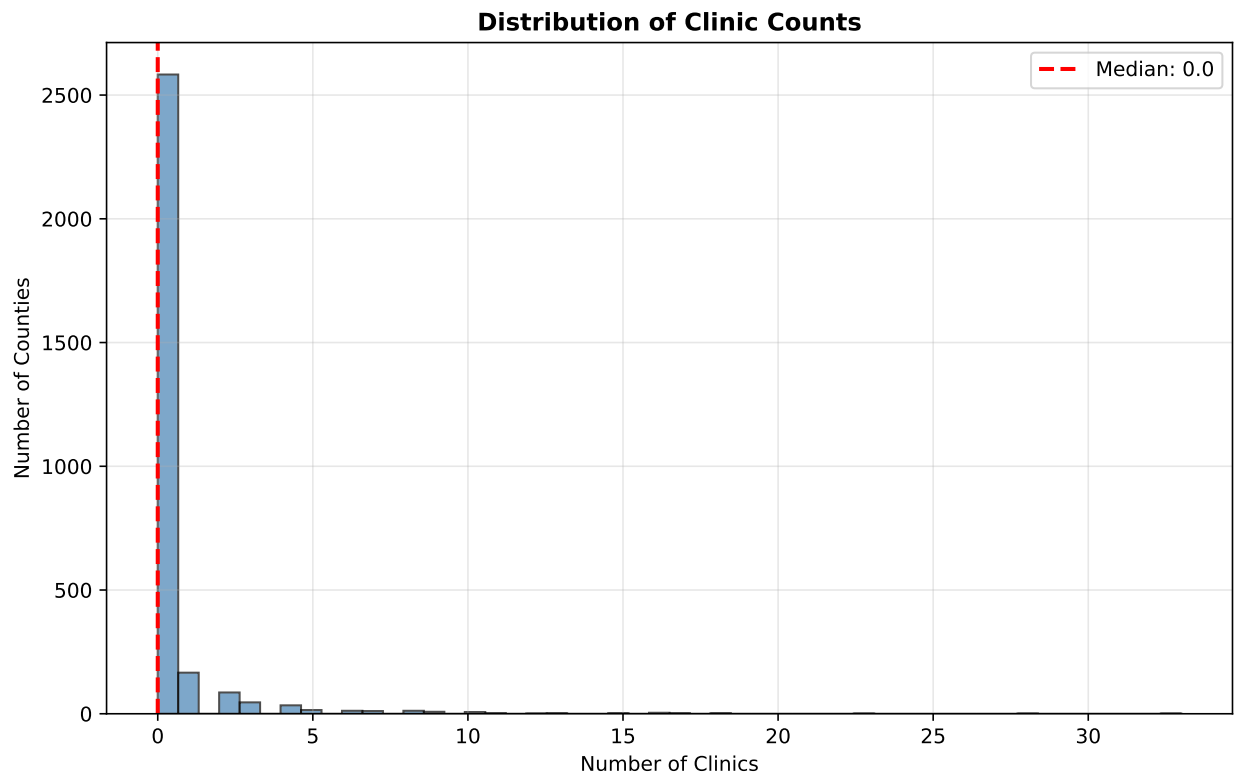


Figure 1: Distribution of CVS clinic counts across all U.S. counties. Most counties have zero clinics.

**what this shows:** the histogram reveals that most counties have zero clinics. the distribution is highly skewed, with a few counties having many clinics while the vast majority have none. this indicates CVS clinics are highly concentrated in specific areas.

## Health Burden Distribution

```
plt.figure(figsize=(10, 6))
plt.hist(df['health_burden_score'], bins=50, edgecolor='black', alpha=0.7, color='purple')
plt.axvline(df['health_burden_score'].mean(), color='red', linestyle='--', linewidth=2,
            label=f'Mean: {df["health_burden_score"].mean():.2f}')
plt.axvline(df['health_burden_score'].median(), color='orange', linestyle='--', linewidth=2,
            label=f'Median: {df["health_burden_score"].median():.2f}')
plt.xlabel('Health Burden Score')
plt.ylabel('Number of Counties')
plt.title('Distribution of Health Burden Scores', fontweight='bold')
plt.legend()
plt.grid(alpha=0.3)
plt.show()
```

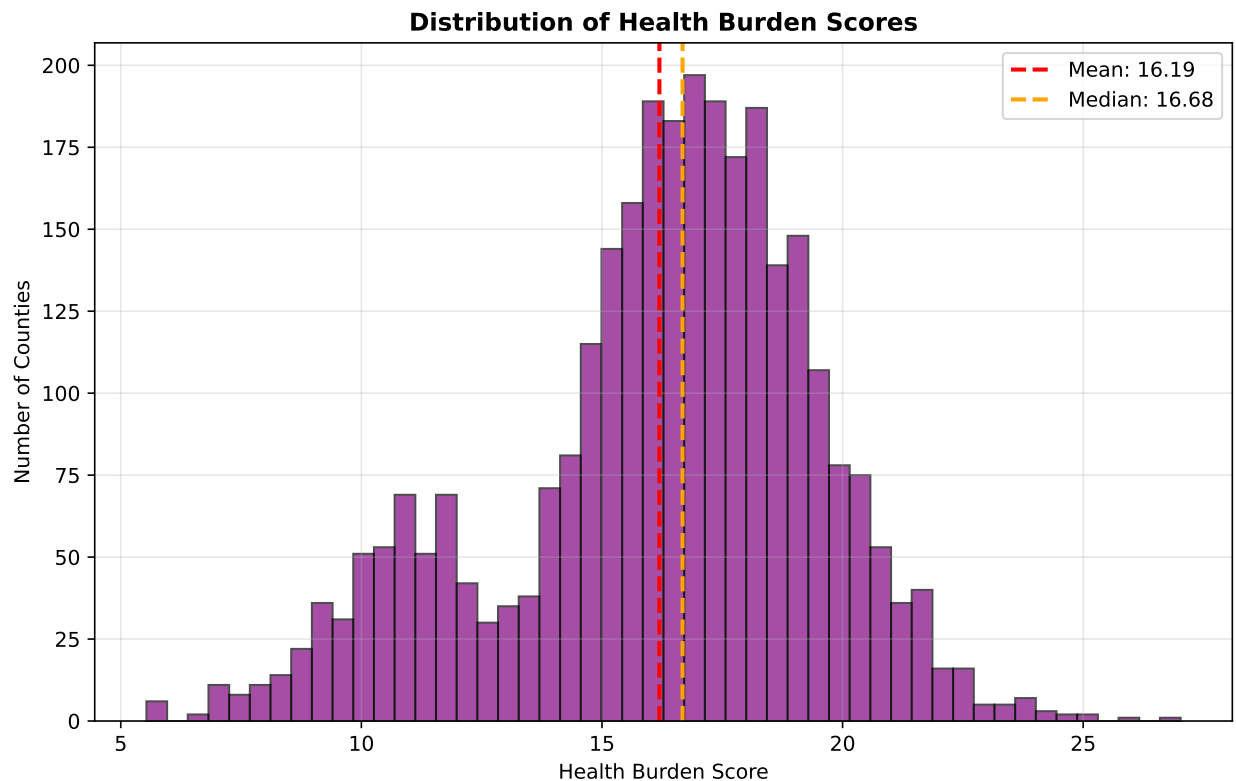


Figure 2: Distribution of health burden scores. Higher scores indicate worse health outcomes.

**what this shows:** health burden scores are normally distributed with some variation. counties with higher scores have worse health outcomes (more stroke, disability, inactivity, social isolation). comparing this to clinic distribution helps identify underserved areas.

## Clinic Presence vs Health Burden

```
df['clinic_presence'] = df['clinic_count'].apply(lambda x: 'No Clinics' if x == 0 else ('1-2 Clinics' if x < 3 else '3+ Clinics'))
comparison_data = df.groupby('clinic_presence')['health_burden_score'].mean()

plt.figure(figsize=(10, 6))
plt.bar(comparison_data.index, comparison_data.values, color=['#d62728', '#ff7f0e', '#2ca02c'])
plt.xlabel('Clinic Presence Category')
plt.ylabel('Average Health Burden Score')
plt.title('Health Burden by Clinic Presence', fontweight='bold')
plt.xticks(rotation=45)
plt.grid(axis='y', alpha=0.3)
plt.show()
```

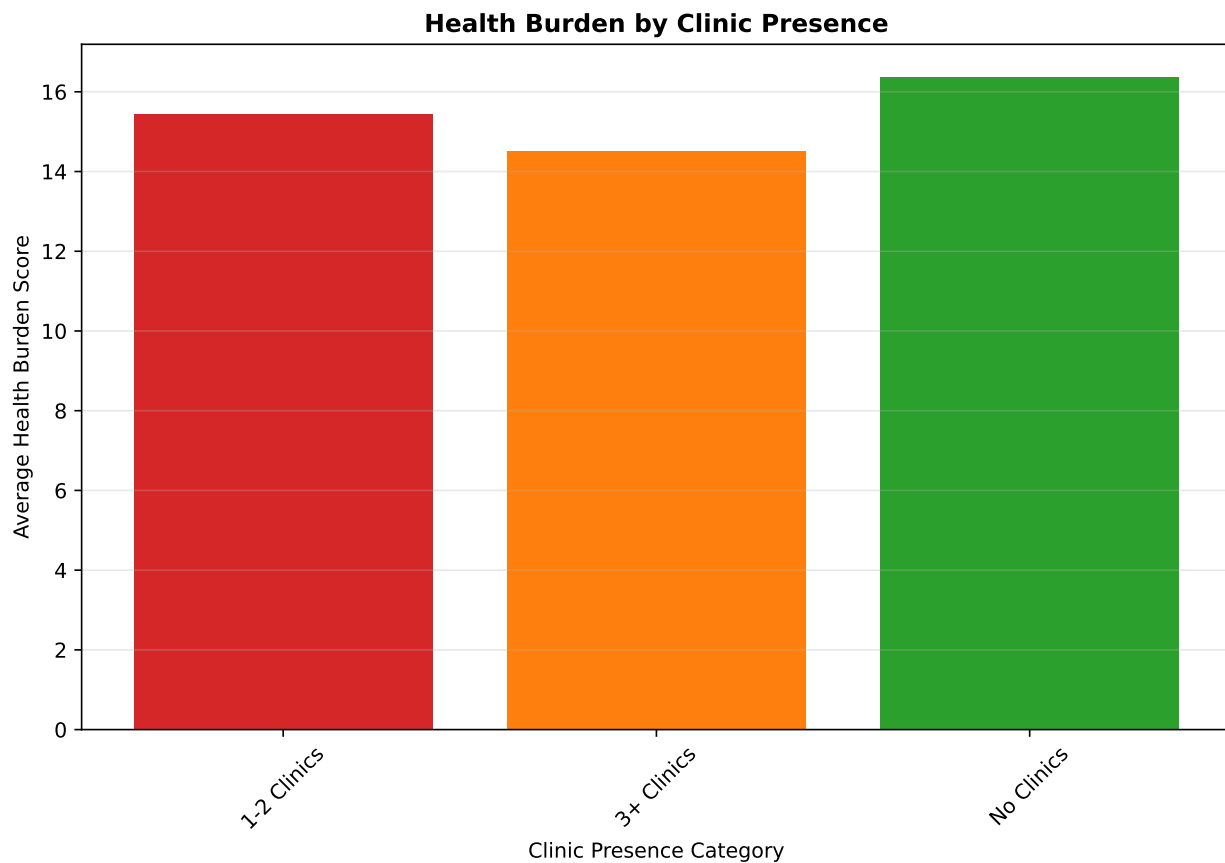


Figure 3: Comparison of health burden scores by clinic presence category.

**what this shows:** counties with no clinics have the highest average health burden (red bar), indicating sicker populations have less access. counties with more clinics tend to have lower health burden, suggesting either better access leads to better health, or clinics are located in healthier areas.

## SVI Quartiles and Clinic Distribution

```
df['svi_quartile'] = pd.qcut(df['svi_overall'], q=4, labels=['Low SVI (Q1)', 'Medium-Low SVI (Q2)', 'Medium-High SVI (Q3)', 'High SVI (Q4)'])
```

```
plt.figure(figsize=(12, 6))
sns.boxplot(data=df, x='svi_quartile', y='clinic_count', palette='viridis')
plt.xlabel('SVI Quartile')
plt.ylabel('Number of Clinics')
plt.title('Clinic Count Distribution by Social Vulnerability Index Quartiles', fontweight='bold')
plt.xticks(rotation=45)
plt.grid(axis='y', alpha=0.3)
plt.tight_layout()
plt.show()
```

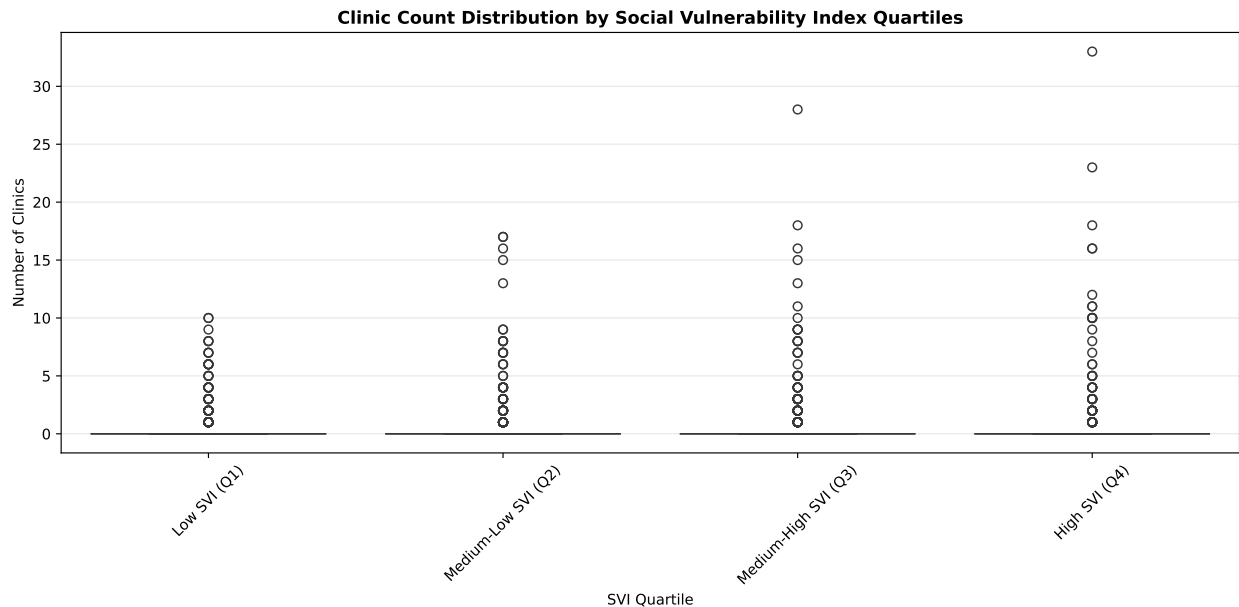


Figure 4: Box plot showing clinic count distribution across SVI quartiles.

**what this shows:** the box plots reveal that counties in higher SVI quartiles (more vulnerable) tend to have fewer clinics. the median clinic count is zero across all quartiles, but the upper quartiles (outliers) show that some low-vulnerability counties have many clinics, while high-vulnerability counties rarely have clinics.

## Geographic Expansion Opportunities Map

this map shows where CVS should expand by combining health need, social vulnerability, and current clinic access. darker colors indicate higher priority expansion targets.

```
# create expansion priority score
# combines health need, SVI, and lack of clinics
# higher score = higher priority for expansion
df['expansion_priority'] = (
    df['health_need'] * 0.4 + # 40% weight on health need
    df['svi_overall'] * 0.3 + # 30% weight on vulnerability
    (1 - df['clinic_availability']) * 0.3 # 30% weight on lack of clinics (inverted)
)

# ensure fips codes are properly formatted
df['fips'] = df['fips'].astype(str).str.zfill(5)
```

```

# download county boundary data for mapping
try:
    url = "https://raw.githubusercontent.com/plotly/datasets/master/geojson-counties-fips.json"
    counties_geo = requests.get(url, timeout=30).json()
    geo_loaded = True
    print("geographic data loaded successfully")
except Exception as e:
    print(f"could not load geographic data: {e}")
    geo_loaded = False

if geo_loaded:
    # create choropleth map showing expansion priority
    fig = px.choropleth(
        df,
        geojson=counties_geo,
        locations='fips',
        color='expansion_priority',
        color_continuous_scale="Reds", # red = high priority
        scope="usa",
        labels={'expansion_priority': 'Expansion Priority Score'},
        title="CVS Expansion Priority Map: Health Need + Vulnerability + Access Gap",
        hover_data=['county_full', 'state_full', 'health_burden_score', 'svi_overall', 'clinic_count']
    )

    fig.update_geos(fitbounds="locations", visible=False)
    fig.update_layout(height=600, title_font_size=16)
    fig.show()

    print("\n expansion priority map created")
    print("  darker red = higher priority for expansion")
    print("  priority based on: health need (40%), SVI (30%), lack of clinics (30%)")
    print(f"  priority score range: {df['expansion_priority'].min():.3f} to {df['expansion_priority'].max():.3f}")
else:
    print("geographic data not available - cannot create map")

```

geographic data loaded successfully

Unable to display output for mime type(s): text/html

Expansion priority map showing counties by health need, SVI, and clinic access.

Unable to display output for mime type(s): text/html

```

expansion priority map created
darker red = higher priority for expansion
priority based on: health need (40%), SVI (30%), lack of clinics (30%)
priority score range: 0.259 to 0.970

```

what this map shows:

- **darker red areas:** counties with high expansion priority - these combine high health need, high vulnerability, and low clinic access
- **lighter areas:** counties with lower priority - may have clinics already, lower health needs, or lower vulnerability

- **hover over counties:** see specific values for health burden, SVI, and clinic count
- **expansion targets:** focus on the darkest red areas where CVS can make the biggest impact
- **geographic patterns:** the map reveals regional clusters of underserved communities (mississippi delta, rural south, etc.)

this map helps visually identify where CVS should prioritize expansion efforts to maximize both business impact and community health outcomes.

## Strategic Recommendations

### Priority 1: High-Impact Urban Expansion

- Focus on large metropolitan counties with high health burden and moderate-to-low clinic density
- Target: Los Angeles County, Harris County (Houston), Dallas County, Maricopa County (Phoenix)
- Rationale: Maximum population impact with existing infrastructure

### Priority 2: Rural High-Need Markets

- Target rural counties with high health burden scores and zero clinics
- Focus on Mississippi Delta, rural South, and Native American communities
- Consider mobile clinics or partnerships with existing healthcare facilities

### Priority 3: Vulnerable Community Access

- Prioritize counties with high SVI scores and zero clinics
- Address socioeconomic barriers to healthcare access
- Consider sliding scale pricing or community health partnerships

### Priority 4: Cluster-Based Expansion

- Use K-means clustering results to identify similar counties
- Develop standardized expansion strategies for each cluster type
- Leverage successful clinic models from similar county types

## Next Steps

1. **Validate Findings:** Cross-reference identified underserved counties with local healthcare infrastructure and competitor presence
2. **Market Research:** Conduct feasibility studies for top-priority expansion targets
3. **Partnership Opportunities:** Explore partnerships with local health systems in underserved areas
4. **Pilot Programs:** Launch pilot clinics in 2-3 high-priority counties to test expansion model
5. **Monitor Impact:** Track health outcomes and utilization rates in new clinic locations
6. **Iterate Strategy:** Use clustering results to refine expansion criteria and identify new opportunities