Safe Drinking Water for California Water Systems*

At Risk Small and Disadvantaged Communities

Susan Peck

October 8, 2025

1 Abstract

This paper looks at whether an environmental, health, and socioeconomic score is related to the risk of small water systems failing in California. Using 2024 data from the Drinking Water Needs Assessment and the CalEnviroScreen tool, the study focuses on systems serving fewer than 3,300 connections. A simple linear regression model tested how the CalEnviroScreen score, which measures pollution burden and population characteristic, relates to the Total Risk Score for each water system. The results show a small but significant positive relationship. The communities with higher pollution tend to have water systems at greater risk. However, the model explains only a small part of the variation, so it should not be used alone to predict risk. The findings suggest pollution data can help guide support for disadvantaged communities but need to be combined with other measures.

2 Introduction

California drinking water is safe for 98% (California State Water Resources Control Board 2024) of the population and meets California state drinking water standards that are more strict than federal regulations. This still leaves about 400 failing water systems serving 870,000 people, 600 water systems serving 1.6 million people that are at risk of failure, and more than 400 others serving another 1.6 million that are potentially at risk of failing. (California State Water Resources Control Board 2025) Smaller communities and populations that are economically disadvantaged may be more at risk of having water with higher levels of contaminants, or they may not have the resources to address and fix problems with the water system.

*Project repository available at: https://github.com/susanpeck/MATH261A-project-1-Linear-Regression

In 2016 the California State Water Resource Board adopted a Human Right to Water Resolution that includes a statement that "every human being has the right to safe, clean, affordable, and accessible water adequate for human consumption, cooking and sanitary purposes." The goals of Human Right to Water are to provide safe drinking water, accessible drinking water, affordable drinking water, and/or maintaining a sustainable and resilient water system. In 2019 California established the Safe and Affordable Funding for Equity and Resilience (SAFER) Program with the goal of helping struggling water systems and to help provide affordable safe drinking water. In 2021 SAFER performed a Needs Assessment and provided recommendations to try to determine where and how funds should be used to have the most impact in improving failing or at risk water systems.

The California State Water Board conducted the risk assessment with 19 indicators across four categories. The categories are Water Quality, Accessibility, Affordability, and Technical, Managerial, and Financial (TMF) Capacity. Each year the Needs Assessment is updated and the data published with risk assessment scores for each subcategory. This paper will focus on 2024 data and the total weighted risk assessment scores from communities with fewer than 3300 service connections.

Funding has been provided to try and address inequities in small water systems and fix problems in a sustainable way. This paper will investigate to see if there are correlations between the water system risk assessment score from the Needs Assessment and a pollution burden score called CalEnviroScreen score calculated by the Office of Health Hazard Assessment. Can we see a connection between populations that have a general higher burden from pollution and water systems that are failing or at risk of failing? Could we use the CalEnviroScreen score to predict communities that should receive help and funding to their water system?

This paper will provide details and definitions for the data obtained from the Drinking Water Needs Assessment, the indicators that are used to calculate the Total Risk for a water system, and the indicators that are used to calculate the CalEnviroScreen Score in the Data Section 3. It will describe the general simple linear regression model and define the variables used to analyze potential correlations in the Methods Section 4. A simple linear regression model will be fitted to the data and show a positive linear association between the CalEnviroScreen Score and the Total Risk Score in the Results Section 5, along with analysis of the residuals. The Discussion Section 6 will talk about the positive association seen in the linear regression model and that though there seems to be some significance to values in the model, it would probably not be good to use the model to predict which communities could use funding based only on the CalEnviroScreen score.

3 Data

This paper uses data from the 2024 Drinking Water Needs Assessment report (California State Water Resources Control Board 2024) through the SAFER program and definitions from the original 2021 Drinking Water Needs Assessment report (California State Water Resources

Control Board 2021). The data was gathered from water systems' annual reports, water quality testing results sent to the California Water Board, and Census data. The reports were prepared by the California State Water Resources Control Board within the California Environmental Protection Agency (CalEPA), in partnership with the UCLA Luskin Center for Innovation (UCLA).

The report calculated a risk assessment value for each water system. The Risk Assessment Result is based off of a score for each of 19 risk indicators (see Figure 1). A standardized score is a value between 0 and 1. Weight values between 1 and 3 were applied to the individual risk indicators. This resulted in a sub-score value for each subcategory. A total weighted risk value, called Total Risk Score for this paper, was calculated using the four subcategories scores. The result was used to indicate if the water system was At-Risk, Not-At-Risk, or Potentially-At-Risk. Some systems were not evaluated and have a Not Assessed value.

Risk Indicator Category	Risk Indicators
Water Quality	History of E. coli Presence
	Increasing Presence of Water Quality Trends Toward MCL
	Treatment Technique Violations
	Past Presence on the HR2W List
	Maximum Duration of High Potential Exposure (HPE)
	Percentage of Sources Exceeding an MCL
Accessibility	Number of Sources
	Absence of Interties
	Water Source Types
	DWR – Drought & Water Shortage Risk Assessment Results
	Critically Overdrafted Groundwater Basin
Affordability	Percent of Median Household Income (%MHI)
	Extreme Water Bill
	% Shut-Offs
TMF Capacity	Number of Service Connections
	Operator Certification Violations
	Monitoring and Reporting Violations
	Significant Deficiencies
	Extensive Treatment Installed

Figure 1: Risk Indicator Categories

All "At Risk" systems exceed a 'threshold of concern' (see Section 8 for specific values) for at least four risk indicators.

The Needs Assessment spent time trying to figure out a threshold for what would make water unaffordable to the residents of that water system. Affordability in this data set is based on three measured values: the percent of the MHI, a comparison to the state average water bill, and the number of shut-offs of water. (Section 8)

Accessibility is a combination of five indicators that include how many sources a water system has, if the system is reliant on connections to other systems, and what types of sources (groundwater, snow, rivers, etc.) This score also includes information about droughts and overdrafted water basins.

TMF Capacity risk indicators measure a system's technical, managerial and financial (TMF) capacity to actually maintain a water system long term.

This paper looks at how the Total Risk Score varies compared to the California Environmental Screening Score for that community. The Total Risk Score is calculated based on weighted sub-scores from Water Quality, Affordability, Accessibility, and TMF Capacity, (See longer definitions in Section 7). The CalEnviroScreen Score is calculated based of two sub-scores from Pollution Burden and Population Characteristics.

The CalEnviroScreen ("CalEnviroScreen" 2025) is a score created by the State of California Office of Environmental Health Hazard Assessment (OEHHA) that uses environmental, health, and socioeconomic information to produce scores for every census tract in the state. The 20 indicators (Section 9) fall into four sub-categories: Exposures, Environmental Effects, Sensitive Populations, and Socioeconomic factors. Each indicator is based on current data from a public agency and every piece of data has a location where the measurement was taken. A higher score is an area that has a higher pollution burden than a lower score.

4 Methods

This paper looks at the results of a simple linear regression model shown below.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Y_i represents the response variable. In this paper the response variable is the Total Risk Score.

X_i represents the predictor variable. In this paper the predictor variable is the CalEnviroScreen Score.

Both the response and predictor variable are scaled and weighted scores based on percentages, so the variables do not have units.

β_0 is the y-intercept of the model. For the linear regression analysis this would be the expected mean value of the Total Risk Score for a small water system with a CalEnviroScreen Score of

0. The theoretical range for the CalEnviroScreen Score is 0 to 100, but in a practical sense 0 is not a possible score.

β_1 is the slope of the model. The slope represents how much the response variable changes for a unit change in the predictor variable. In the regression model the slope would represent how much we would expect the mean Total Risk Score to change based on a one unit change for the CalEnviroScreen Score.

ε_i are the error terms.

This model assumes a linear relationship between Y and X. If the actual relationship between Y and X is not a linear one, then the model could be a bad fit for the data, to the extreme of not being useful at all, or possibly just have some bias and predictive values are off.

The model also assumes that the error terms are independent, have a constant variance for all levels of the predictor variable X, the mean of the error terms is zero, and the error terms are normally distributed. If these assumptions are not true then the values and confidence intervals can be misleading. For this data set the sample size is large enough that the normality assumption is less important.

The analysis was done using R programming {R Core Team (2025)} and built in function, lm, to fit the linear model and provide calculated estimate values for the slope and intercept of the model. The summary function of the lm model also shows a p-value for each of the estimates and a R^2 value.

5 Results

After a couple years of funding and some improvements to small water systems in California, are there correlations between risk indicator scores? To begin this analysis this paper looked at a scatterplot of the CalEnviroScreen Score and MHI (Figure 2). Only smaller water systems with fewer than 3300 service connections are included. Each data point is colored to show if the water system is at risk of failing to provide clean drinking water by separating the data into four risk categories.The larger the CalEnviroScreen Score, the more pollution a population may be subject to. General pollution may or may not directly impact water quality. The scatterplot below indicates lower MHI seems to have more At-Risk water systems. The larger the value of the MHI for the water system, the fewer water systems that have high CalEnviroScreen score. This scatterplot is included to show the general range of CalEnviroScreen scores for the small water systems in this paper and possible grouping of risk levels.

A second scatterplot (Figure 3) again graphs CalEnviroScreen Score against MHI, but this time groups the data by the economic status of the service area. The status could be a Disadvantaged Community (DAC), a Severely Disadvantaged Community (SDAC), or a Non-Disadvantaged Community (Non-DAC). The classifications of how disadvantaged a community is is a direct calculation based on MHI, as shown by the vertical sections on the scatterplot.

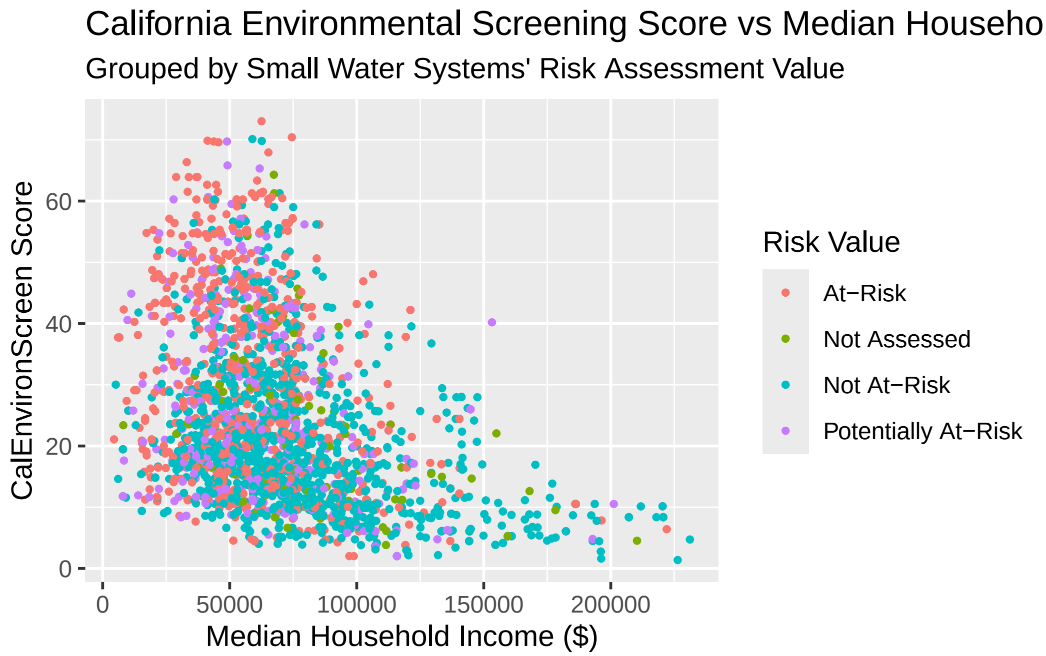

Figure 2: Scatter plot of calenvironscreen versus MHI grouped by risk assessment values

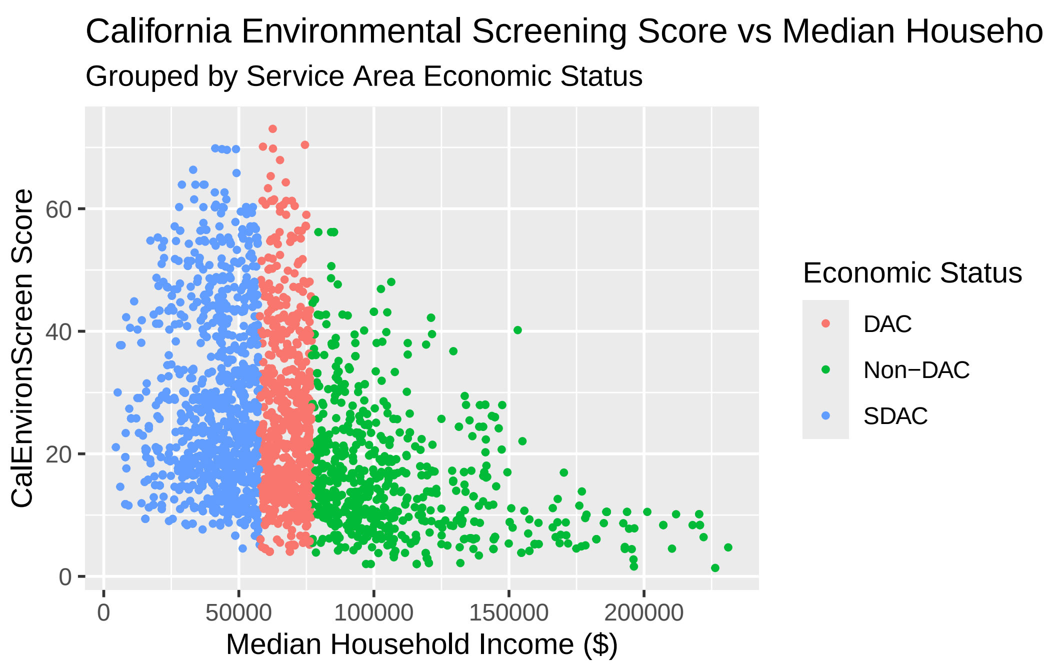

Figure 3: Scatter plot of calenvironscreen versus MHI grouped by service area economic status

Both the DAC and SDAC communities have a large range of CalEnviroScreen Scores. The Non-DAC communities have a smaller range of values and the values are less in general.

The SDAC communities CalEnviroScreen Score ranged from 4.551 to 69.86 with and average value of 28.688. The DAC communities CalEnviroScreen Score ranged from 4.035 to 73.036 with and average value of 24.858.The Non-DAC communities CalEnviroScreen Score ranged from 1.372 to 56.183 with and average value of 16.389.

Now this paper will look at how the Total Risk Score varies with CalEnviroScreen Score, and if a linear model is a good choice for the association. In Figure 4 a scatterplot is created showing a possible positive association between Total Risk Score and CalEnviroScreen Score. Included in the scatterplot is a linear regression line.

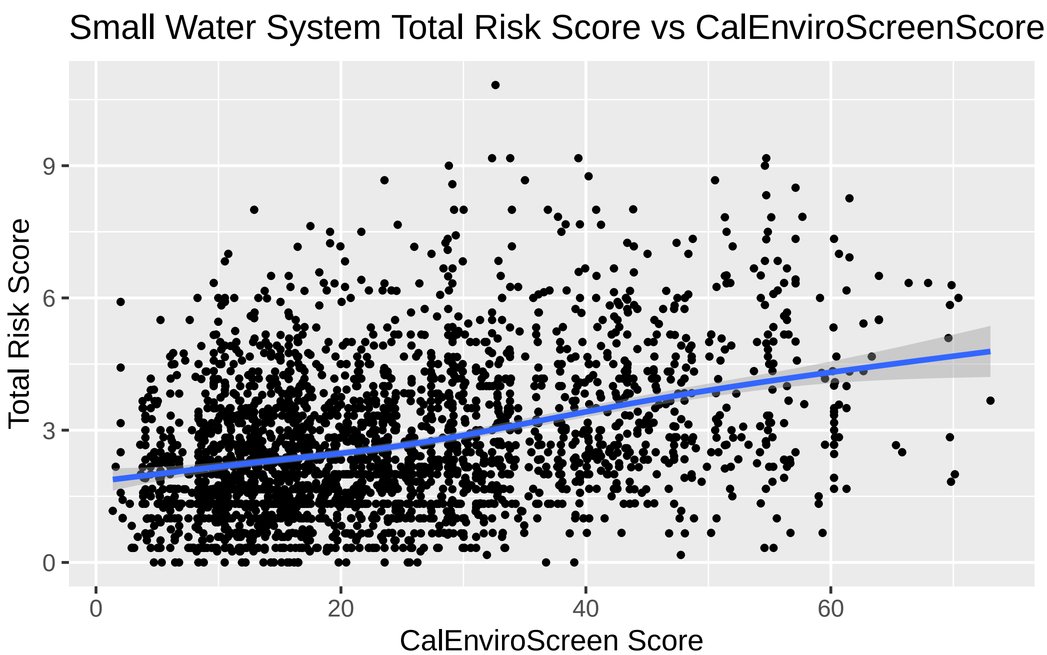

Figure 4: Scatter plot of CalEnvironScreen Score versus Total Risk Score

The simple linear regression model for this relationship would have a b_0 value of 1.685 and the b_1 value of 0.0425. As discussed in the Methods Section 4, the intercept b_0 does not have great interpretation because a predictor value of CalEnviroScreen Score equal to zero is not a reasonable value for an actual data point. The b_1 value represents the slope of the linear model and means that for every unit increase in the CalEnviroScreen Score for a small water system, the Total Risk Score has increased by about 0.0425. The positive slope matches the visual on the scatterplot that there is a positive association.

In the model the p-value is small (much smaller than 0.001) for both the intercept and slope estimates, indicating there may be some significance to the model. However, the large range of values in the data for each predictor level, and the smaller R^2 value of 0.1272761 means that the model may not be the best fit for the data and may not do a good job of explaining the variance.

The CalEnviroScreen score is the result of several factors that may contribute to pollution in an area and is not explained purely by the income level of the local population.

Below is a residual plot from the linear regression model, plotting the residual against the fitted value. If this plot shows random points above and below zero with no clear trend then this suggests that the model could be a good fit for the data. If the residuals show a trend that is curved or something that is not horizontal at y=0, it could suggest some of the assumptions in the model are not met.

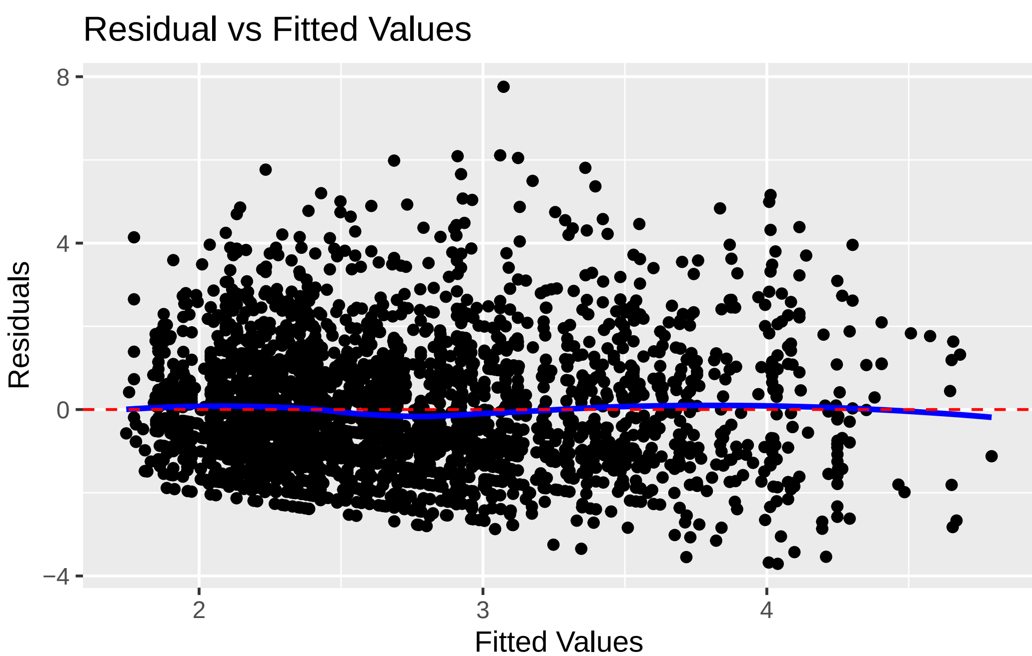

Figure 5: Residual plot of Total Risk Score versus CalEnvironScreen Score

The residual plot Figure 5 has a small trend line, dipping down in the middle and is not perfectly horizontal at y=0, but there is not a large trend in the residuals suggesting a significant pattern that is not explained by the model. The residuals appear to be spread wider above the y=0 line, possibly suggesting that it is not a symmetrical distribution. The residuals also have a wider range on the right of the plot, fanning out slightly, meaning the assumption of constant variance in the error terms is not met. The p-test or t-test values may not be reliable.

In the Q-Q plot of the residuals Figure 6 the values stray far from the y=x line, suggesting that the assumption of normality in the errors is not met. However, this analysis has over

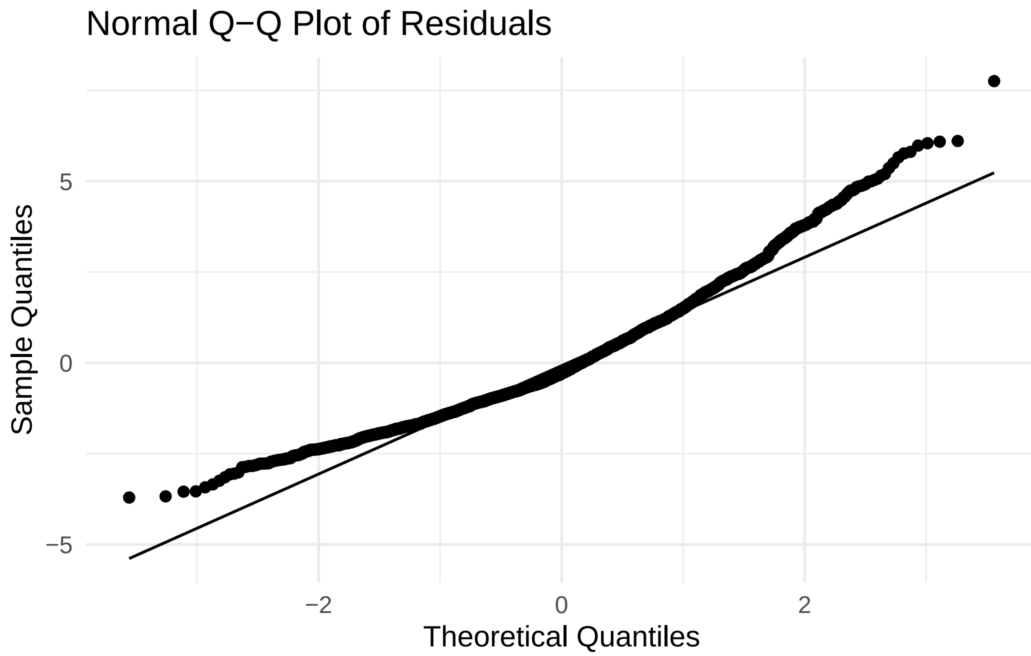

Figure 6: Q-Q plot for Residuals of Total Risk Score versus CalEnvironScreen Score Linear Regression

2000 data points which limits the reliance on the assumption of normality and does not mean the model is useless or useful by itself.

6 Discussion

If the CalEnviroScreen Scores are a good predictor of the Total Risk Score response variable, a possible application of the model would be to use the CalEnviroScreen scores to decide which small water systems get funding and resources. Right now there are multiple agencies looking at similar data and ideas: can we identify factors put water quality and water systems at risk for contamination or failing to provide safe and accessible water. This puts a strain on resources. If we can use a score or other information that already exists to predict with some certainty which water systems are at risk of failing for any reason, then the funding can have a more direct path.

The linear model in this paper suggests that there is a positive association between the two scores, and a linear model is an appropriate model to use. However, both scores are based on a weighted calculation of approximately 20 indicators each. And some of the indicators in both calculations overlap in the information they provide. Both scores look at socioeconomic status of the populations served in different locations and use it as part of the calculation. This is a positive because if a similar indicator is a good predictor of whether a water system is at risk of failing, then one score could be used in place of the other. However, that overlap limits how much we can trust the predictive value of the linear regression model because that association could be based on that similar indicator.

There is still work to be done to evaluate whether or not the funding and improvements are actually helping the small water systems that are at risk of failure. We still do not know if the Total Risk Score is correctly identifying the communities with the highest need or highest risk, and if the improvements are making sustainable long term improvements.

7 Appendix A DEFINITIONS

Affordability Threshold means the level, point, or value that delineates if a water system's residential customer charges, designed to ensure the water systems can provide drinking water that meets State and Federal standards, are unaffordable. For the purposes of the 2021 Affordability Assessment, the State Water Board employed affordability thresholds for the following indicators: Percent Median Household Income; Extreme Water Bill; and Percent Shut-Offs. Learn more about current and future indicators and affordability thresholds in Appendix E.

Affordability Assessment means the identification of any community water system that serves a disadvantaged community that must charge fees that exceed the affordability threshold established by the State Water Board in order to supply, treat, and distribute potable water that complies with Federal and state drinking water standards. The Affordability Assessment evaluates several different affordability indicators to identify communities that may be experiencing affordability challenges. (Health & Saf. Code, § 116769, subd. (2)(B).

At-Risk public water systems or *At-Risk PWS* means community water systems with 3,300 service connections or less and K-12 schools that are at risk of failing to meet one or more key Human Right to Water goals: (1) providing safe drinking water; (2) accessible drinking water; (3) affordable drinking water; and/or (4) maintaining a sustainable water system.

Median household income or *MHI* means the household income that represents the median or middle value for the community. The methods utilized for calculating median household income are included in Appendix A and Appendix E. Median household incomes in this document are estimated values for the purposes of this statewide assessment. Median household income for determination of funding eligibility is completed on a system by system basis by the State Water Board's Division of Financial Assistance.

Risk indicator means the quantifiable measurements of key data points that allow the State Water Board to assess the potential for a community water system or a transient noncommunity water system that serves a K-12 school to fail to sustainably provide an adequate supply of safe drinking water due to water quality, water accessibility, affordability, institutional, and/or TMF capacity issues.

Risk threshold means the levels, points, or values associated with an individual risk indicator that delineates when a water system is more at-risk of failing, typically based on regulatory requirements or industry standards.

Score means a standardized numerical value that is scaled between 0 and 1 for risk points across risk indicators. Standardized scores enable the evaluation and comparison of risk indicators.

Service connection means the point of connection between the customer's piping or constructed conveyance, and the water system's meter, service pipe, or constructed conveyance, with certain exceptions set out in the definition in the Health and Safety Code. (See Health & Saf. Code, § 116275, subd. (s).)

Small community water system means a CWS that serves no more than 3,300 service connections or a yearlong population of no more than 10,000 persons. (Health & Saf. Code, § 116275, subd. (z).)

Risk Indicators means quantifiable measurements of key data points that allow the State Water Board to assess the probability of a water system's failure to deliver safe drinking water or other infrastructure and institutional failures. Risk indicators that measure water quality, accessibility, affordability, and TMF capacity are incorporated based on their criticality as it relates to a system's ability to remain in compliance with safe drinking water standards and their data availability and quality across the State.

Risk Indicator Thresholds are the levels, points, or values associated with an individual risk indicator that delineates when a water system is more at-risk of failing.

Scores & Weights are the application of a multiplying value or weight to each risk indicator and risk category, as certain risk indicators and categories may be deemed more critical than others and/or some may be out of the control of the water system. The application of weights to risk indicators and risk categories allows the State Water Board multiple ways to assess all risk indicators within each category together in a combined Risk Assessment score.

TMF Capacity risk indicators measure a system's technical, managerial and financial (TMF) capacity to plan for, achieve, and maintain long term compliance with drinking water standards, thereby ensuring the quality and adequacy of the water supply.

8 Appendix B THRESHOLD VALUES

Risk Indicator	Thresholds	Score	Weight
History of E. coli Presence	**Threshold 0** = **No** history of E. coli presence within the last three years.	0	N/A
	Threshold 1 = **Yes**, history of E. coli presence (E. coli violation and/or Level 2 Assessment) within the last three years.	1	3
Increasing Presence of Water Quality Trends Toward MCL	**Threshold 0** = **No** Increasing Presence of Water Quality Trends Toward MCL.	0	N/A
	Threshold 1 = **Secondary Contaminants**: 9-year average of running annual average is at or greater than 80% of MCL and running annual average has increased by 20% or more.	0.25	2
	Threshold 2 = **Primary Non-Acute Contaminants**: 9-year average of running annual average is at or greater than 80% of MCL and running annual average has increased by 5% or more.	0.5	2
	Threshold 3 = **Acute Contaminants**: • 9-year average (no running annual average) is at or greater than 80% of MCL; or • 24-month average is at or greater than 80% of MCL; or • Any one sample over the MCL.	1	2
Treatment Technique Violations	**Threshold 0** = **0** Treatment Technique violations over the last three years.	0	N/A
	Threshold 1 = **1 or more** Treatment Technique violations over the last three years.	1	1

Risk Indicator	Thresholds	Score	Weight
Past Presence on the HR2W List	**Threshold 0** = **0** HR2W list occurrence over the last three years.	0	N/A
	Threshold 1 = **1** HR2W list occurrence over the last three years.	0.5	2
	Threshold 2 = **2 or more** HR2W list occurrences over the last three years.	1	2
Maximum Duration of High Potential Exposure (HPE)	**Threshold 0** = **0** years of HPE over the last nine years.	0	N/A
	Threshold 1 = **1** year of HPE over the last nine years.	0.25	3
	Threshold 2 = **2** years of HPE over the last nine years.	0.5	3
	Threshold 3 = **3 or more** years of HPE over the last nine years.	1	3
Percentage of Sources Exceeding an MCL	**Threshold 0** = **less than 49.9%** of sources exceed an MCL.	0	N/A
	Threshold 1 = **50% or greater** of sources exceed an MCL.	1	3
Number of Sources	**Threshold X** = **0** sources.	Automatically At-Risk	N/A
	Threshold 0 = **multiple** sources.	0	N/A
	Threshold 1 = **1** source only.	1	3
Absence of Interties	**Threshold 0** = **1 or more** interties.	0	N/A
	Threshold 1 = **0** interties.	1	1
Water Source Types	**Threshold 0** = **2 or more** water source types.	0	N/A
	Threshold 1 = **1** water source type and that source is **purchased** water.	0.5	1
	Threshold 2 = **1** water source type and that source is either **groundwater or surface water**.	1	1

Risk Indicator	Thresholds	Score	Weight
DWR – Drought & Water Shortage Risk Assessment Results	**Threshold 0** = **Below top 25%** of systems most at risk of drought and water shortage.	0	N/A
	Threshold 1 = **Between top 25% - 10.01%** of systems most at risk of drought and water shortage.	0.25	2
	Threshold 2 = **Top 10%** of systems most at risk of drought and water shortage.	1	2
Critically Overdrafted Groundwater Basin	**Threshold 0** = **Less than 74.99%** of system's service area boundary is within a critically overdrafted basin.	0	N/A
	Threshold 1 = **75% or greater** of systems service area boundary is within a critically overdrafted basin.	1	2
Percent of Median Household Income (%MHI)	**Threshold 0** = **Less than 1.49%**	0	N/A
	Threshold 1 = **1.5% - 2.49%**	0.75	3
	Threshold 2 = **2.5% or greater**	1	3
Extreme Water Bill	**Threshold 0** = **Below 149.99%** of the statewide average.	0	N/A
	Threshold 1 = **150% - 199.99%** of the statewide average.	0.5	1
	Threshold 2 = **Greater than 200%** of the statewide average.	1	1
% Shut-Offs	**Threshold 0** = **less than 9.99%** customer shut-offs over the last calendar year.	0	N/A
	Threshold 1 = **10% or greater** customer shut-offs over the last calendar year.	1	2
Number of Service Connections	**Threshold 0** = **greater than 501** service connections.	0	N/A
	Threshold 1 = **500 or less** service connections.	1	1

Risk Indicator	Thresholds	Score	Weight
Operator Certification Violations	**Threshold 0** = **0** Operator Certification violations over the last three years.	0	N/A
	Threshold 1 = **1 or more** Operator Certification violations over the last three years.	1	3
Monitoring & Reporting Violations	**Threshold 0** = **1 or less** Monitoring & Reporting violations over the last three years.	0	2
	Threshold 1 = **2 or more** Monitoring & Reporting violations over the last three years.	1	2
Significant Deficiencies	**Threshold 0** = **0** Significant Deficiencies over the last three years.	0	N/A
	Threshold 1 = **1 or more** Significant Deficiencies over the last three years.	1	3
Extensive Treatment Installed	**Threshold 0** = **No** extensive treatment installed.	0	N/A
	Threshold 1 = **Yes**, extensive treatment is installed.	1	2

9 Appendix C CalEnviroScreen Score Indicators

Group	Subtype / sub-group	Indicators	Brief description
Pollution Burden	*Exposure Indicators*	Ozone	Average daily maximum ozone concentration
		PM .	Fine particulate matter concentrations (using monitors + satellite data)
		Diesel Particulate Matter	Diesel PM emissions (on-road, off-road, and transboundary)
		Traffic Impacts	Volume of traffic (vehicle miles traveled) near census tract / road proximity
		Pesticide Use	Quantity of agricultural pesticide applied in or near the tract
		Toxic Releases from Facilities	Amount of toxic chemical releases from industrial facilities (e.g. TRI)
		Drinking Water Contaminants	Concentrations / violations of contaminants in drinking water systems serving the tract
		Children's Lead Risk from Housing	Risk of lead exposure to children from older housing stock with lead paint (especially in low-income areas)
	Environmental Effects Indicators	Cleanup Sites	Presence / density of hazardous site cleanup efforts or polluted sites
		Groundwater Threats	Threats to groundwater quality (e.g. from agricultural practices, contamination)
		Hazardous Waste Generators and Facilities	Density or proximity of hazardous waste facilities / generators
		Impaired Water Bodies	Watersheds or water bodies failing quality standards that affect the area
		Solid Waste Sites and Facilities	Landfills, transfer stations, illegal dumping sites near the tract

Group	Subtype / sub-group	Indicators	Brief description
Population Characteristics	*Sensitive Population Indicators*	Asthma	Prevalence or rates of asthma in the population
		Cardiovascular Disease	Rates of heart disease / cardiovascular conditions in the population
		Low Birth Weight Infants	Rates of infants born of low birth weight (a vulnerability indicator)
	Socioeconomic Factor Indicators	Educational Attainment	Percent (or share) of population with low education levels (e.g. < high school)
		Housing Burden	Share of households with high housing cost burdens (e.g. > some % of income) ([Ceja][5])
		Linguistic Isolation	Percentage of households in which no one age 14 speaks English "very well"
		Poverty	Fraction of population (or households) below poverty thresholds
		Unemployment	Unemployment rate in the tract or labor force without jobs

References

"CalEnviroScreen." 2025. State of California Office of Environmental Health Hazard Assessment (OEHHA). https://oehha.ca.gov/calenviroscreen.

California State Water Resources Control Board. 2021. "2021 Drinking Water Needs Assessment Executive Summary." https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/executive_summary.pdf.

———. 2024. "2024 Drinking Water Needs Assessment Results." https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/2024/2024-needs-assessment.pdf.

———. 2025. "2025 SAFER Dashboard." https://www.waterboards.ca.gov/safer/.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.