Safe Drinking Water for California Water Systems*

At Risk Small and Disadvantaged Communities

Susan Peck

September 24, 2025

Abstract

This will be a great abstract summarizing everything succinctly and also making sure that people want to read the rest of my paper. It will be intriguing and informative. I understand that not having an abstract in my draft means that you can not give me constructive and timely feedback on this part of my paper.

Introduction

California drinking water is safe for 98% (California State Water Resources Control Board 2024) of the population and meets 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.

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

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

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 to try to determine where and how funds should be used to have the most impact in improving failing or at risk water systems and provided recommendations.

The California State Water Board conducted the risk assessment with 19 indicators across four categories. The categories were Water Quality, Accessibility, Affordability, and Technical, Managerial, and Financial (TMF) Capacity. Each year the Needs Assessment was updated and the data published with risk assessment scores for each category. This paper will focus on 2024 data and the scores from each of the categories 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 economic status of the population a small water system serves and indicator scores.

This paper will detail the data obtained from the Drinking Water Needs Assessment and relevant variables, describe the general simple linear regression model used to analyze potential correlations, and show the results of the linear regression and residual analysis.

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 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 below). 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 score value for each subcategory. A total weighted risk value 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

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

This paper looks at the individual subcategory scores from the Risk Assessment, a California Environmental Screening Score, population, and Median Household Income (MHI), and investigates the possible relationships between those variables. The subcategories are Water Quality, Affordability, Accessibility, and TMF Capacity. (See longer definitions from the Needs Assessment report Section)

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. A higher score is an area that has a higher pollution burden than a lower score. Below is a scatterplot of the CalEnvironScreen score compared to MHI for that water system. 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.

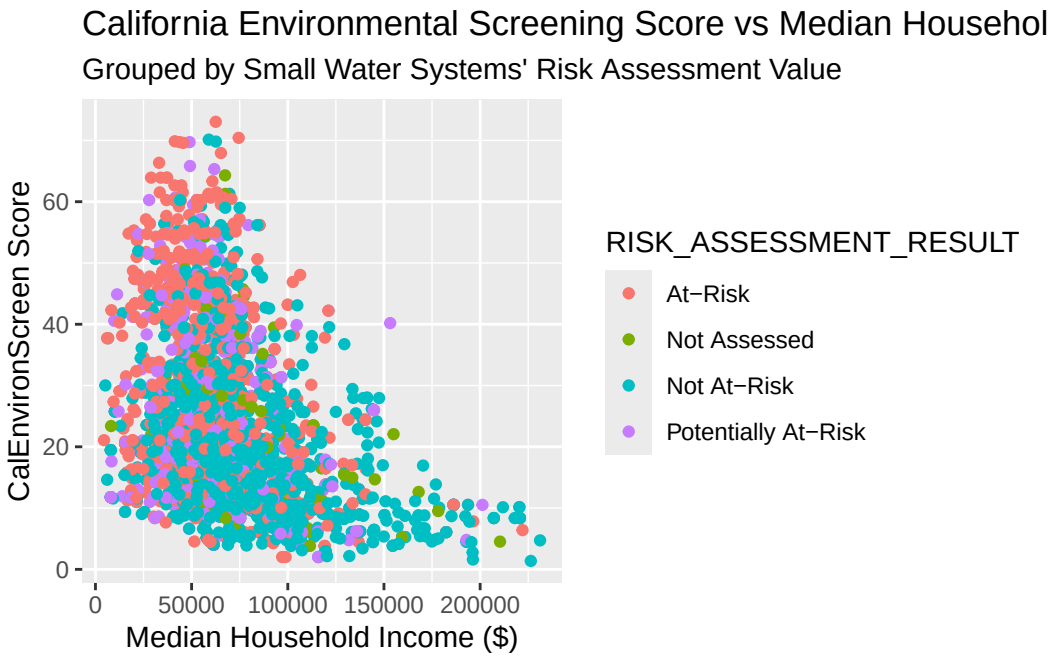

Figure 1: Scatter plot of calenvironscreen versus MHI grouped by risk assessment values

The Median Household Income (MHI) uses the annual income value for all the people in a single household. The MHI for this data set is the median household annual income for the population in each water system. The median household income is a typical value that is used as a general economic measure and as a way to compare populations. On an individual level, MHI can be used to determine if you qualify for certain aid. In this data set MHI is one way to compare different populations served by different water systems.

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

Accessibility is a combination of five indicators that include how many sources a water system has, if the system is reliant on connetions 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.

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 first linear regression analysis the response variable is the CalEnviroScreen Score.

X_i represents the predictor variable. In this paper the first linear regression analysis the predictor variable is MHI.

β_0 is the y-intercept of the model. For the first linear regression analysis this would be the expected mean value of the CalEnviroScreen score for water system population with a median household income of \$0. A household could have an income of \$0, but it is unlikely to be a median household income for a population, so this particular intercept doesn't have a direct interpretation.

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 first regression model the slope would represent how much we would expect the mean CalEnviroScreen Score to change based on a one dollar change for the MHI.

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.

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 we looked at a scatterplot of the CalEnviroScreen Score and MHI Figure 1 shown in the Data Section Section of this paper. The larger the CalEnviroScreen Score, the more pollution a population is 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.

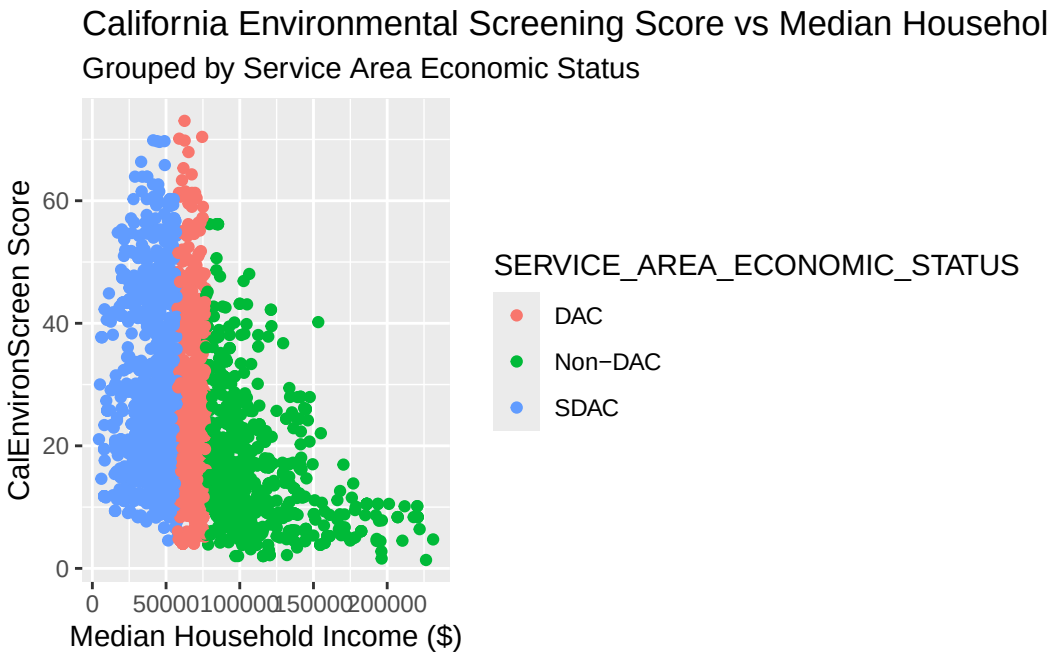

Figure 2: Scatter plot of calenvironscreen versus MHI grouped by service area economic status

The scatterplot above again graphs CalEnviroScreen Score against MHI, but this time groups 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. 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.

The simple linear regression model for this relationship would have a b_0 value of 34.546789 and the b_1 value of $-1.5345561 \times 10^{-4}$.

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.

Apendix 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

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/drinking_water/certlic/drinkingwater/documents/needs/2024/2024-needs-assessment.pdf.

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