Correlation Between Water Quality and Mortality in California

Vincent Zheng, Uzi Garcia, Nolan Johnson

Abstract

We study the association between drinking-water contaminants and all-cause mortality in California using a county–year panel from 2000–2019. We estimate log-linear Poisson models with county and year fixed effects and an offset for population to obtain incidence rate ratios (IRRs) for each analyte. Across contaminants, most IRRs are statistically indistinguishable from 1, indicating limited evidence of strong associations at this spatial–temporal resolution. Occasional small effects appear but are not robust across specifications and should not be interpreted causally given observational data, potential exposure mismeasurement, and unobserved time-varying confounders. The analysis highlights which contaminants merit closer study and provides a reproducible template for linking water-quality monitoring data to population health outcomes.

Introduction

Access to clean drinking water is a foundational determinant of public health, and contamination by naturally occurring or anthropogenic chemicals has been linked to a variety of adverse health outcomes. In California, home to over 39 million people and characterized by diverse geology, agriculture, and urban development, both water quality and mortality rates exhibit substantial geographic and temporal variation. Additionally, California is one of the few states with thorough tracking of water quality data, including concentrations of regulated contaminants across public water systems. Yet few studies have systematically quantified how county-level fluctuations in specific contaminants relate to all-cause mortality controlling for unobserved local and temporal factors.

In this analysis, we assemble a county–year panel for California by merging two administrative sources over 2000-2019. Annual mortality counts (Y_{ct}) and population totals (N_{ct}) are drawn from the California Department of Public Health's "Death Profiles by ZIP Code" dataset, then aggregated up to the county level[1]. Water contaminant concentrations[2] (X_{ct}) come from the State Water Resources Control Board's Contaminant Data Browser, which provides measurements for a comprehensive list of regulated analytes (e.g., arsenic, vanadium, radon) across public water systems. By aligning these data at the county–year scale, we capture average exposures alongside

[1] California Department of Public Health, *Death Profiles by ZIP Code* [data set], California Open Data Portal, accessed July 30, 2025, https://data.ca.gov/dataset/death-profiles-by-zip-code.

[2] State Water Resources Control Board, *Contaminant Data Browser* [data set], accessed July 30, 2025, https://www.waterboards.ca.gov/resources/data_databases/contaminants.html.

corresponding mortality outcomes.

Model specification

Let

$$Y_{ct} \sim \text{Poisson}(\lambda_{ct}) \tag{1}$$

denote the number of deaths in county c and year t. We link the mean λ_{ct} to contaminant exposure X_{ct} via a log-linear regression with county and year fixed effects, offset by population:

$$\log(\lambda_{ct}) = \beta_0 + \beta_1 X_{ct} + \alpha_c + \gamma_t + \log(N_{ct}) \tag{2}$$

- β_0 is the intercept.
- β_1 captures the change in log–mortality rate per one-unit increase in the contaminant concentration.
- α_c absorbs all time-invariant county characteristics.
- γ_t absorbs state-wide shocks common to year t.
- The offset $\log(N_{ct})$ converts death counts into rates per person.

This Poisson-fixed-effects framework allows us to estimate the incidence rate ratio $\exp(\beta_1)$, i.e. the multiplicative change in the mortality rate associated with a one-unit rise in exposure, while controlling for both space and time confounders.

Results

Table 1 presents the incidence rate ratios (IRRs) from (2). All models include county and year fixed effects and an offset for the log population.

A one-unit (µg/L) increase in contaminant concentration is associated with a multiplicative change in the mortality rate equal to the IRR. Equivalently, $(\text{IRR} - 1) \times 100\%$ gives the percent change in the mortality rate per unit increase.

For example, a one-µg/L increase in **combined uranium** is associated with a 0.005% decline in the mortality rate. In contrast, most other analytes yield IRRs indistinguishable from 1 at conventional significance levels. This example also cautions against interpreting the IRR as a causal effect, since uranium in drinking water shouldn't decrease mortality.

All estimates are drawn from observational data, so unobserved time-varying confounders (such as changes in treatment practices or socioeconomic conditions) may drive part of the observed relationships. Nonetheless, our results highlight which water-quality measures exhibit the strongest links to mortality rates and can guide further, more targeted investigation.

Table 1: Percent Changes of Incidence Rate Ratios for Water Contaminants

Analyte	% Change	p-value	Signif
aggressive index	-0.001%	<0.001	***
alkalinity, bicarbonate	+0.000%	0.925	
alkalinity, total	-0.000%	0.296	
arsenic	-0.001%	0.376	
boron, total	-0.000%	0.287	
bromide	-0.011%	0.105	
bromochloroacetic acid	-0.019%	0.257	
calcium	-0.000%	0.741	
carbon dioxide	-0.000%	0.714	
carbon, total	+0.002%	0.002	**
chlorate	-0.000%	0.033	*
chloride	+0.000%	0.195	
chromium, hex	-0.001%	0.077	
combined uranium	-0.005%	0.015	*
conductivity @ 25 c umhos/cm	+0.000%	0.029	*
dibromoacetic acid	-0.027%	0.030	*
dichloroacetic acid	-0.012%	0.058	
fluoride	+0.024%	0.004	**
gross alpha particle activity	-0.002%	0.148	
gross beta particle activity	+0.004%	0.296	
hardness, total (as caco3)	+0.000%	0.097	
langelier index (ph(s))	+0.043%	0.115	
langelier index at source temp.	-0.000%	<0.001	***
magnesium	+0.000%	0.178	
manganese, dissolved	+0.002%	<0.001	***
molybdenum, total	+0.010%	0.493	
nitrate	-0.002%	0.260	
nitrate-nitrite	-0.000%	0.035	*
orthophosphate	+0.001%	0.357	
ph	-0.001%	0.001	**
ph, field	+0.000%	0.375	
phosphate, total	+0.002%	0.013	*
potassium	+0.001%	0.104	
radon	-0.000%	0.020	*
silica	+0.000%	0.448	
sodium	+0.000%	0.454	
sodium adsorption ratio	-0.000%	0.044	*
sulfate	+0.000%	0.401	
tds	+0.000%	0.079	
temperature (centigrade)	+0.000%	0.922	
total haloacetic acids (haa5)	-0.001%	0.536	
total radium for ntnc per 64442(b)(3)	+0.100%	0.141	
trichloroacetic acid	-0.001%	0.865	
tthm	-0.001%	0.042	*

Table 1: Percent Changes of Incidence Rate Ratios for Water Contaminants *(continued)*

Analyte	% Change	p-value	Signif
turbidity	+0.001%	0.219	
turbidity, field	+0.000%	0.912	
vanadium, total	+0.008%	<0.001	***

Significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$