



Appendix: COVID-19 Impact on Household Water System Debt at Zip Code Level

Analysis of State Water Board Survey Data by UCLA Luskin Center for Innovation

Gregory Pierce, Peter Roquemore, Jacqueline Adams, Angelica Ruiz and Silvia R. González
(contact: gspierce@ucla.edu)

March 2021

Background

This document provides further details regarding the UCLA Luskin Center for Innovation's (UCLA) support to the State Water Board to produce the Board's analysis of the Drinking Water COVID-19 Financial Impacts Survey, which was publicly released and presented on January 19, 2021. UCLA carried out this work as a final task completed under the AB 401 contract with the Board.

More detail regarding the Board's work analyzing the survey and the survey data can be found for public use at the following webpage:

https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/covid-19watersystemsurvey.html.

UCLA researchers provided support to the Board by a) validating some aspects of the water system financial analyses, b) independently analyzing and assisting with the validation of zip code-level results presented on January 19, 2021, and c) generating the zip code level debt findings focused on socio-demographics of debt concentration. The below description summarizes UCLA's supporting methods and findings for the zip code demographic analysis.

Brief Summary of Data Processing and Analysis

Two separate surveys were sent out by the Board: one to community water systems with fewer than 10,000 service connections, and one to community water systems with 10,000 service connections or more. In turn, systems were asked in the surveys to report both system-level and zip code level customer debt estimates. Among returned surveys, data quality was fairly high, leading to exclusion of relatively few erroneous entries or questionable outliers.

Survey data also provided debt figures at the zip code level, reported by participating community water systems, for 1051 of California's 1763 zip codes. Zip codes in the data contain 75- 80% of the state's population and indicate there are at least 1.2 million delinquent accounts, before projecting figures statewide. Zip codes reporting debt data were not notably different from those areas not reporting debt, except that the reporting areas tended to be slightly more urban, and thus dense and diverse.

Debt data at the zip code level was provided in debt “buckets” in \$100 increments from “Under \$100,” “\$100 - \$200,” up to “Over \$1000.” Each bucket was assigned a number of accounts in that zip code that hold a debt level within that bucket. To calculate the monetary value of the debt, each bucket is reassigned a midpoint value and multiplied by the number of accounts that fall within the bucket. For example, if there are 100 accounts in a “\$100 - \$200” bucket, debt is calculated by multiplying 100 * \$150 for a total of \$15,000, and this process is continued for each debt bucket per zip code. Additionally, all debt over \$1000 was conservatively assumed to be \$1050. This assumption likely skews overall projections low. There are approximately 150,000 accounts with over \$1000 in debt, which could contribute a considerable amount of debt to statewide totals.

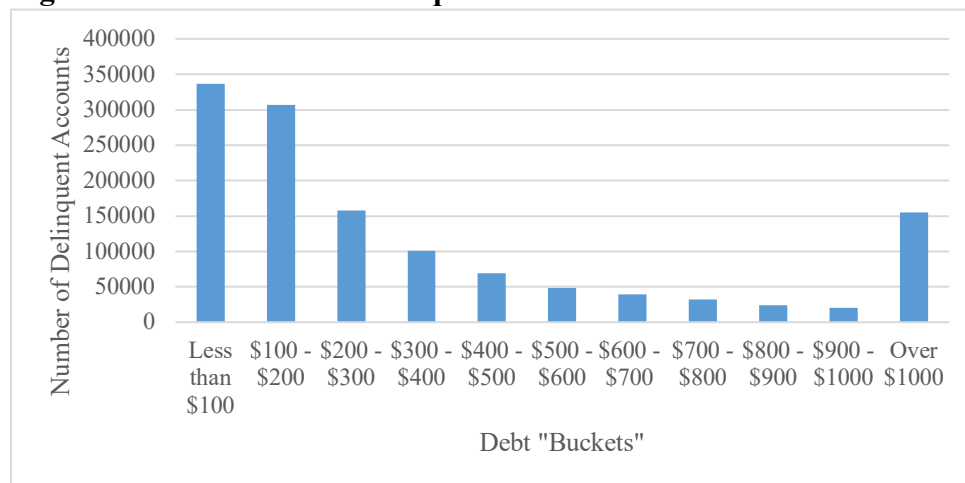
Descriptive, bivariate correlational and multivariate regression analysis methods of the survey data points were then employed by UCLA researchers. These methods facilitated the calculation of statewide debt prevalence and level estimations, the identification of zip codes with high prevalence and levels of debt, and the estimation of relationships between debt and socio-economic status factors at the zip code level.

Statewide projections for debt were analyzed alongside socio-demographic and socio-economic status using zip code level 2018 American Community Survey data for population, race, economic, and housing characteristics, provided by the UCLA Center for Neighborhood Knowledge, and joined to zip code level debt estimates by zip code ID (ZCTA). Analysis of debt trend relationships to socio-demographic factors was enabled by matching American Community Survey zip code tabulation area (ZCTA) data to corresponding zip codes. Multivariate regression analysis of relationships between debt and other zip code characteristics (n=808) was performed in the statistical analysis software program Stata 16.1.

Key Zip Code Level Findings

Our analysis yielded the following key findings linking debt to socio-demographic and socio-economic factors. As Figure 1 shows, debt levels tend to be concentrated at the low and high ends of the measured distribution.

Figure 1. Distribution of Delinquent Accounts across Debt Levels



Debt levels, however, were not equally distributed across zip codes. Rather, concentration of debt was found in zip codes with higher percentages of racial and ethnic minority residents. In fact, having a higher percentage of Hispanic and Black households in a zip code is statistically-significantly, positively associated with multiple measures of greater debt burden.

Zip codes with high percentages of Hispanic and Black households were found to also have:

- A higher percentage of households with some level of debt,
- A higher average level of debt,
- A higher percentage of households with very high levels of debt: above \$600 and \$1,000, and
- These relationships hold true even after controlling for income, poverty level and housing type.

In zip codes that are predominately Hispanic or Black, the majority of debt is concentrated in Disadvantaged Community (“DAC”) and Severely-Disadvantaged Community (“SDAC”) zip codes.

- Overall, 47% of household debt comes from DAC zip codes. However, in zip codes with over 60% Hispanic or Black populations, those that are also DAC represent 77% and 97% of debt respectively, and
- In predominately Hispanic or Black zip codes, the average debt per delinquent household increases as average incomes decrease.

Multivariate Regressions

Multivariate regression further aided in assessing which, if any, socio-demographic and socio-economic factors were most strongly associated with debt. Three overarching questions were examined, which, if any, socio-demographic and socio-economic factors may help explain:

- The % of households in a zip code with any level of debt?
- The average level of debt in a zip code?
- The % of households in a zip code with very high levels of debt?

Factors assessed at the zip code level for their relationship to debt prevalence and magnitude in each multivariate regression were:¹

- Race/ethnicity: the % of the zip code’s households reported as Non-Hispanic White, Hispanic, Black, and Asian
- Economic status: median household income & % of households in poverty
- Housing status: the % of households living in rental units.

In brief, using both Ordinary Least Squares (OLM) and Generalized Linear Model (GLM) types, we found the following relationships between different and socioeconomic factors. A “+” sig

¹ We also considered other socio-demographic variables available from ACS data including citizenship status, health insurance status, and vehicle ownership status. However, we excluded these variables from our final set of regressions for several reasons: missing or suppressed data for some zip codes, and collinearity concerns with included variables combined with the lesser conceptual relevance of variables to the outcome of interest.

indicates a positive, statistically significant relationship and the (#) indicates the relative influence of the factor in explaining the debt outcome in each regression, with “1” being the most influential factor.

Table 1. Reduced Form Multivariate Regression Model Results of Zip Code Level Debt Outcomes

Debt outcome modeled	% of households in debt	Average debt level	Average debt level across all households above \$100	% of households above \$600 debt	% of households above \$1000 debt
% Black Households	+ (2)	+ (2)	+ (1)	+ (2)	+ (2)
% Hispanic households	+ (1)	+ (1)	+ (3)	+ (1)	+ (1)
Median income				+ (4)	+ (4)
% households in poverty					
% of households renting housing unit	+ (3)	+ (3)	+ (2)	+ (3)	+ (3)
Total variation in outcome explained (Adjusted R ²)	10%	12%	10%	9%	7%

Predominantly Minority Zip Codes

Finally, we also examined debt levels in zip codes with predominantly minority populations and disadvantaged community (DAC) or severely disadvantaged community (SDAC) status. We define predominantly minority zip-codes as where Black, Latinix, or Asian residents make up 60% or more of the population. The threshold is informed by research from the UCLA Center for Neighborhood Knowledge’s COVID-related research on high vulnerability zip codes.² DAC and SDAC status are defined by a zip code’s average median household income (MHI) in relation to the statewide average. DAC zip codes are those with an MHI at 80% or below of the statewide MHI and SDAC zip codes have an MHI at 60% or below the statewide average.

² Latinx, Black, and some Asian American populations are heavily concentrated in neighborhoods designated as either high or highest levels of medical vulnerability (72%, 70%, 69%, respectively). See Ong, P., Pech, C., Rios. Gutierrez, N., Mays, V. (2020). “COVID-19 Vulnerability Indicators: California Data for Equity in Public Health Decision Making.” UCLA Center for Neighborhood Knowledge and UCLA BRITE Center for Science, Research and Policy.

Overall, 47% of household debt comes from DAC zip codes. However, in zip codes with over 60% Hispanic or Black populations, those that are also DAC represent 77% and 97% of debt respectively. As Table 2 shows, in predominately Hispanic or Black zip codes, the average debt per delinquent household increases as the median income of the area decreases.

Table 2. Zip Code Debt per Delinquent Household in Predominantly Minority Zip Codes

	All Zip Codes	Non-Hispanic White	Hispanic	Black	Asian
Total Debt per Delinquent HH	\$362.48	\$380.86	\$377.76	\$485.52	\$249.27
DAC Debt per Delinquent HH	\$366.68	\$292.12	\$386.57	\$496.99	\$234.75
SDAC Debt per Delinquent HH	\$374.96	\$204.67	\$397.87	\$422.85	NA