Aesthetics Level Occurrence Study

Executive Summary

Principal Investigator:

Carleigh Samson, Corona Environmental Consulting, LLC

Submitted to Water Quality Research Foundation on August 17, 2020

Overview

Drinking water treatment processes are designed to both reduce health risks caused by exposure to contaminants, as well as reduce aesthetic issues, such as color, taste, and odor in drinking water. Aesthetic issues are often the most concerning drinking water issues for many consumers because they are easily and immediately identifiable. In addition to setting regulations for health-related drinking water contaminants, the US Environmental Protection Agency (USEPA) has established Secondary Drinking Water Standards (USEPA, 2020a), which are non-mandatory water quality standards for aesthetic analytes known to impact the color, taste, and odor of drinking water. The non-enforceable standards, referred to as secondary maximum contaminant levels (SMCLs), are set to give public water systems (PWSs) guidance on removing chemicals causing aesthetic issues to levels that are below what most people will find to be noticeable. These secondary standards include iron, manganese, and pH, among other analytes. In addition to the analytes included in the secondary standards, chlorine and chloramine residual and hardness are important water quality factors that impact the aesthetics of drinking water.

The occurrence of aesthetic analytes in drinking water and their negative impact on color, taste, and odor can cause concern and dissatisfaction among consumers. The application of point-of-use (POU) and point-of-entry (POE) devices can provide an opportunity, beyond centralized treatment, to maintain a high standard of water quality through the removal of contaminants causing aesthetic issues.

The objectives of this project were to:

- compile a national database of drinking water quality occurrence data for selected aesthetic analytes impacting the color, taste, and odor of drinking water
- assess the concentration and frequency of, and population affected by, the occurrence of selected aesthetic analytes

This report summarizes the national occurrence of aesthetic analytes that impact the color, taste, and odor of drinking water. These analytes include chlorine/chloramine residual, hardness, iron, manganese, and pH.

Approach

Data Collection

The data collection effort focused on collecting all available drinking water occurrence data for the ten-year period from the start of 2009 through the end of 2018 for as many PWSs as possible across the US. A data outreach effort was conducted to request drinking water occurrence data from regulatory agencies for all 50 states in the US. Outreach to all primacy agencies for the 50 US states was completed in the second half of July 2019. The entire state data collection effort spanned more than 7 months, with the last state's data set received in early March 2020. The data request provided to contacts at the primacy agencies asked for all drinking water occurrence data records for public water systems for, at minimum, the period from the start of 2009 through the end of 2018.

The USEPA's UCMR4 data collection includes data for 30 unregulated contaminants collected in 2018 through 2020. Data for manganese samples collected in 2018 from the USEPA's UCMR4 were downloaded from the USEPA's website (https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule#4). UCMR4 data collected after 2018 were not included at this point for consistency with the project database, which houses data from January 1, 2009 through December 31, 2018. Manganese data from PWSs in 50 US states, the District of Columbia, 8 tribal regions, and 4 territories were available for these UCMR4 analytes.

The USEPA's federal SDWIS Water System Summary and Water System Details reports for the fourth quarter of 2019 were downloaded for all PWSs from the USEPA federal SDWIS website (https://ofmpub.epa.gov/apex/sfdw/f?p=108:1:0::NO:1). From these reports, specific system characteristics were collected for use in the database, including system type, system size based on population served, and primary source water type.

Data QA/QC

A rigorous quality assurance and quality check (QA/QC) process was conducted on all data records collected from the state data collection effort and from the UCMR4 data set. As received, the data sets from different states and the UCMR4 data set include different data fields and naming conventions for the data fields. To create one database for all the collected data, the first step of the QA/QC process was to establish a consistent set of data fields for each data set.

State data sets used different analyte naming conventions and analyte codes, which also required standardizing to develop one database for all data records. One outcome of the QA/QC process was to develop a master list of analyte names and codes for all analytes collected in the data collection process and a crosswalk between the various analyte names and codes used by individual state and UCMR4 data sets and the master list of analyte names and codes. The master list and the crosswalk were updated with each additional data set that underwent the QA/QC process.

The QA/QC process also standardized the unit of measure for each project analyte. For each analyte, a standard unit of measure was chosen and a list of acceptable units of measure were determined, such that data could be converted from the acceptable units of measure to the standard unit of measure.

A range of plausible results were developed for each analyte. Statistical analyses of data outliers, a review of USEPA's Six Year Review Three (SYR3) outcomes, and research into common detection limits for standard laboratory analysis methods were explored to set thresholds for in-range and out-of-range data.

All data records were assigned a concentration flag to describe the data and what changes were made during the QA/QC process (i.e. concentration changed due to converting unit of measure). Based on the concentration flag, data records were then assigned a "Retain" or "Discard" flag. Examples of reasons to discard data records include non-sensical units of measure, no unit of measure for a detected concentration, and missing, NA, or null concentration with no indication of whether the data record was below detection or an indication that the data record was not below detection. Data records that were assigned a "Retain" flag and were not out of range were included in the project database.

The QA/QC process did not include a sample location analysis, which would allow for the characterization of data records as either "raw water" or "finished water" samples. Raw water is representative of the source water quality (i.e. groundwater well, river, lake, etc.), and finished water is representative of water quality entering the distribution system or water quality at a customer's tap. Additionally, samples could be collected within the treatment process (i.e. chlorine, pH). The sample location may have an impact on the analysis results, and as a result, data analysis outcomes presented for aesthetic analytes may not be representative of the water quality delivered to consumers. A detailed sample location analysis could provide the opportunity to summarize data specifically for finished water samples.

Database Development

The project database was set up using a cloud-based PostgreSQL (Postgres) database to house all the data collected from state regulatory agencies, UCMR4, and the USEPA federal SDWIS. Postgres is a relational database management system. Once the QA/QC process was completed for each individual state's data records and the UCMR4 data set, they were uploaded to the Postgres database and incorporated into a database table. In total, the database includes data for approximately 68,000 drinking water systems and approximately 12.3 million data records.

The database includes five chlorine/chloramine analytes: "Chloramine", "Chlorine", "Free Residual Chlorine", "Residual Chlorine", and "Total Chlorine" due to different naming and coding conventions on a state by state basis. For states and systems with data for "Chlorine" and "Residual Chlorine", it is unclear whether free or total chlorine was measured. The five separate measures of chlorine/chloramine residual concentration were maintained in the project database to avoid making any assumptions about the disinfectant type.

Data Analysis and Results

The data analysis effort focused on assessing the occurrence of chlorine/chloramine residual, hardness, iron, manganese, and pH, and summarizes the frequency of, and the population affected by, these aesthetic analytes. In the summaries below, national median, or 50th percentile, and 95th percentile values are used to describe the data occurrence. The median or 50th percentile is a measure of central tendency of the data, where 50% of the data are equal to or below the median, and the other 50% are equal to or greater than the median. The 95th percentile represents the extreme highs in the data, where 95% of the data are equal to or below the 95th percentile, and the remaining 5% of the data are greater than or equal to the 95th percentile.

Chlorine/Chloramine Residual

Chlorine/chloramine residual data were available for 20,726 drinking water systems, and the 2009 – 2018 median concentration was equal to 1.0 mg/L and the 95th percentile concentration was equal to 2.0 mg/L. A total of 9,936 community water systems (CWSs), or 86% of CWSs with available data, had a system-wide 95th percentile chlorine/chloramine residual between non-detect and 2.0 mg/L. Collectively, those systems serve a population of more than 60 million. Seventy-nine CWSs, or 0.7%, serving a total population of 828,560, had a system-wide 95th percentile exceeding the maximum residual disinfectant level goal (MRDLG) and maximum residual disinfectant level (MRDL) of 4.0 mg/L.

A total of 5,459 non-transient non-community water systems (NTNCWSs) and transient non-community water systems (TNCWSs), or 86% of NTNCWSs and TNCWSs with available data, which serve a total population of approximately 1.7 million, had a system-wide 95th percentile chlorine/chloramine residual between non-detect and 2.0 mg/L. Ninety-seven NTNCWSs and TNCWSs, or 1.5%, which serve a population of almost 28,000 people, had a 95th percentile chlorine/chloramine residual greater than the MRDL of 4.0 mg/L.

Since 2015, annual 95th percentile chlorine/chloramine residual were found to be highest for very large systems, serving populations greater than 100,000, followed by large systems, serving populations greater than 10,000 and equal to or less than 100,000, compared with systems of other sizes. Comparing systems based on source water type showed that annual 95th percentile chlorine/chloramine residuals were higher in surface water systems and groundwater under the direct influence of surface water systems as compared with groundwater systems.

Hardness

There were total hardness data available for 28,356 systems. The USGS classifies soft water as 0-60 mg/L as $CaCO_3$, moderately hard water as 61-120 mg/L as $CaCO_3$, hard water as 121-180 mg/L as $CaCO_3$, and very hard water as greater than 180 mg/L as $CaCO_3$. Median hardness for all national occurrence data in 2009 – 2018 was 131 mg/L as $CaCO_3$, and the 95th percentile was 490 mg/L as $CaCO_3$. A total of 6,073 CWSs, or 46%, serving a total population of 59 million had a median hardness concentration equal to or greater than 120 mg/L $CaCO_3$. A total of 5,839 NTNCWSs and TNCWSs, or 47%, serving a total population of 1.2 million, had a median hardness concentration equal to or greater than 120 mg/L $CaCO_3$.

Comparing systems based on system size showed that annual 95th percentile hardness concentrations in 2009 – 2018 were the highest for very large systems, serving populations greater than 100,000, and the lowest for medium systems, serving populations greater than 3,300 and less than or equal to 10,000. Overall, CWSs had the highest annual 95th percentile hardness concentrations as compared with NTNCWSs and TNCWSs. Among source water types, surface water systems had the highest annual 95th percentile hardness concentrations, and groundwater under the direct influence of surface water had the lowest annual 95th percentile hardness concentrations. Similar analysis looking the central tendencies of the data as opposed to the extreme highs showed different results, where groundwater under the influence of surface water systems had the highest annual median values followed by surface water systems and then groundwater systems.

Iron

Iron data were available for 48,003 systems. The national iron median concentration for 2009 – 2018 was non-detect and the 95th percentile was 1.6 mg/L. There were 1,835 CWSs, or 7.5%, serving a total population of more than 6 million, with a system-wide median concentration above 0.5 mg/L, exceeding the SMCL of 0.3 mg/L, and 5,047 CWSs, or 21%, serving a total population of more than 27 million, with a system-wide 95th percentile above 0.5 mg/L. There were 2,973 NTNCWs and TNCWSs, or 16%, serving a total population of approximately 574,000, had a system-wide median iron concentration above 0.5 mg/L, and there were 4,881 NTNCWs and TNCWSs , or 26%, serving a total population of approximately 1 million, had a system-wide 95th percentile iron concentration above 0.5 mg/L.

Very small systems, serving populations less than 500, followed by small systems, serving populations greater than 500 and less than or equal to 3,300, had the highest annual 95th percentile iron concentrations compared with larger systems. Very large systems, serving greater than 100,000 people, had the lowest annual 95th percentile iron concentrations. Among system types, TNCWSs had the highest annual 95th percentile iron concentrations, while CWSs had the lowest. Groundwater and groundwater under the direct influence of surface water systems that the highest annual 95th percentile iron concentrations as compared with surface water systems.

Manganese

Data were available for manganese for 2009 – 2018 for 42,866 systems with a national median concentration of non-detect, and the 95th percentile concentration was 336 µg/L. There were 2,908 or 12% of CWSs, serving a total population of approximately 7.3 million, with a system-wide median manganese concentration greater than the SMCL of 50 µg/L, and 5,660 or 23% of CWSs, serving a total population of more than 40 million, with a system-wide 95th percentile manganese concentration greater than the SMCL. There were 2,335 or 17% of NTNCWSs and TNCWSs, serving a total population of approximately 472,000, that had a system-wide median concentration greater than the manganese SMCL of 50 µg/L, and 3,160 or 24% of NTNCWSs and TNCWSs, serving a total population of approximately 715,000, with a system-wide 95th percentile manganese concentration greater than the SMCL.

Very small systems, serving populations less than 500, had the highest annual 95th percentile manganese concentrations compared with systems of larger sizes, and very large systems, serving populations greater than 100,000, and large systems, serving populations greater than 10,000 and less than or equal to 100,000, had the lowest annual 95th percentile manganese concentrations. Among system types, CWSs had the lowest annual 95th percentile manganese concentrations, and surface water systems had the lowest annual 95th percentile manganese concentrations among the source water types.

pH

pH data were available for 32,106 systems in 2009 – 2018. The median pH was equal to 7.5 and the 95th percentile was equal to 8.6. There were 846 or 5% of CWSs with available pH data, serving a total population of 5.2 million, that had a 5th percentile pH less than or equal to 6, and 390 or 2% of CWSs, serving a total population of 14.8 million, that had a 95th percentile pH above 9. There were 11,192 or 67% of CWSs, serving a total population of approximately 108 million, that had a median pH between 7 and 8. There were 829 or 7% of NTNCWSs and TNCWSs, serving a total population of approximately 175,000, that had a 5th percentile pH below 6, and 144 or 1% of NTNCWSs and TNCWSs, serving a total population of approximately 30,000, had a 95th percentile pH above 9. There were 7,468 or 62% of NTNCWSs and TNCWSs, serving a total population of approximately 2 million, that had a median pH between 7 and 8.

The annual 5th and 95th percentile pH values were the lowest for very small systems, serving populations less than 500, while they were the highest for very large systems, serving populations greater than 100,000, and large systems, serving populations greater than 10,000 and less than or equal to 100,000. Among system types, CWSs had the highest annual 5th and 95th percentile pH values, and among source water types, surface water systems typically had the highest 5th and 95th percentile pH values.

Gaps and Opportunities for Additional Research

The effort presented in this report focused on the collection of drinking water quality data for chlorine and chloramine residual, total hardness, iron, manganese, and pH from state regulatory agencies and from the USEPA's UMCR4 data set, the QA/QC of these data, the development of a database to house these data, and a preliminary analysis of the resulting database. The effort leaves research gaps, which present important opportunities for future investigation to better understand the national occurrence of these aesthetic analytes in drinking water provided by PWSs.

A sample location analysis for all data records included in the database is one example of a data analysis gap that was outside of the scope of this effort. The data records collected from state regulatory agencies include important metadata, such as sample type, sample point type, sample point description, and facility type. These data fields provide information to characterize data records by sample location, such as “raw water” or “finished water” samples, where raw water is representative of the source water quality (i.e. groundwater well, river, lake, etc.), and finished water is representative of water quality entering the distribution system or water quality at customers’ taps. At the start of the project, there was not a clear understanding of the challenges in accurately characterizing data records based on their sample location as raw water or finished water samples. The challenges are due to the different approaches among states and systems to identify whether a sample was a raw or finished water sample, or if the sample was collected within the treatment process. In some instances, this information is provided in the sample type data field, while other states use the sample type data field to identify whether a sample was a routine or compliance sample versus a special sample. In other instances, the sample point type data field can be used to identify whether a sample is a raw or finished water sample type. Furthermore, different states and systems use different terminology for sample location. For example, raw water samples could be identified as “Raw”, “RW”, “Raw Water” or by sample point codes such as “RW01”, “RW02”, etc. Due to the different terminologies and different data fields that could be used to express this information, a sample location analysis would need to be conducted on a state by state basis for the 46 states for which data was received in order to best characterize data records as raw or finished water samples. Once a detailed sample location analysis is completed, updates could also be made to the preliminary findings presented in this report to show the same statistical analyses specifically for finished water samples, representative of the water quality at consumers’ taps.

An analysis of systems using free chlorine versus chloramines for secondary disinfection in the distribution system is an important research area that could be investigated with the database developed as part of this project, along with USEPA databases, i.e. UCMR4, UCMR3, and SYR3. The analysis would determine the secondary disinfection practice in use for analytes “Chlorine”, “Chlorine, Residual”, and “Chlorine, Total”, which are not specific to either free chlorine or chloramines.

A co-occurrence analysis of the aesthetic analytes collected as part of this effort is another example of a gap that presents an opportunity for future research. The database provides an enormous amount of national occurrence data that could be used to study the co-occurrence of pH and chlorine/chloramine residual, pH and hardness, pH and iron, and pH and manganese in order to better understand the national co-occurrence of these analytes. Additionally, if the analysis described above to identify systems, and corresponding data records, as either using free chlorine or chloramines, an analysis of relative occurrence of hardness, iron, manganese, and pH could be conducted based on the type of secondary disinfection practice.