Contaminant Level Occurrence Above the MCLG

Executive Summary

Principal Investigator:

Carleigh Samson, Corona Environmental Consulting, LLC

Submitted to Water Quality Research Foundation on August 17, 2020

Overview

The US Environmental Protection Agency (USEPA) sets a regulatory maximum contaminant level (MCL) or action level and a maximum contaminant level goal (MCLG) for all drinking water contaminants included in the National Primary Drinking Water Regulations (NPDWR). The MCLG is the maximum level of a contaminant in drinking water at which no known or anticipated health effect would occur. MCLGs are set based solely on the toxicity of the contaminant and represent an ideal for drinking water standards. The MCL is set as close as possible to the MCLG to protect public health, while considering available technologies and their costs to achieve contaminant removal.

Drinking water contaminants that have MCLs or action levels set at a higher level than the MCLG may occur in public drinking water at levels that may impact public health despite regulatory compliance. This is true for all contaminants with MCLGs of zero, including arsenic, lead, and some disinfection byproducts (DBPs), as drinking water treatment processes and analytical methods cannot reliably ensure complete removal of a contaminant. The potential impact on public health of contaminant occurrence above the MCLG provides an important opportunity for drinking water treatment beyond that provided by public water systems, such as point of use (POU) or point of entry (POE) devices.

The objectives of this project were to:

- compile a national database of drinking water quality occurrence data for contaminants with an MCL or action level greater than the MCLG
- assess the concentration and frequency of, and population affected by, the occurrence of contaminants above the MCLG but below the MCL or action level

This report summarizes the national occurrence of drinking water contaminants at levels that may present potential public health risks despite federal regulatory compliance. The 46 health-based contaminants that will be investigated in this project have regulatory maximum contaminant levels (MCLs) or action levels above their maximum contaminant level goal (MCLG), which represents the maximum level of a contaminant in drinking water at which no known or anticipated health effect would occur. Of the 46 included health-based contaminants, the report specifically focuses on:

- Arsenic
- Lead
- Disinfection byproducts (DBPs)

Background

Drinking water regulations under the USEPA's Safe Drinking Water Act (SDWA) aim to protect public health through means achievable by all public water utilities. To determine the contaminants of concern that should be regulated, the USEPA conducts a monitoring effort every five years under the Unregulated Contaminants Monitoring Rule (UCMR) to determine the occurrence of up to 30 unregulated contaminants. Once the USEPA has identified a new contaminant to regulate, it must evaluate both the toxicity of the contaminant as well as the cost implications of treatment to reduce or remove the contaminant. Based on the toxicity alone, the USEPA will set the MCLG, which represents an ideal for drinking water standards. Available technology or the cost burden to achieve the necessary contaminant removal to meet the MCLG may not be feasible for all public water systems. The USEPA must consider the technologies available and their costs in setting the regulatory MCL. The MCL is set as close to the MCLG as possible to protect public health, while allowing the opportunity for all public water utilities to achieve regulatory compliance.

The USEPA's NPDWRs currently include legally enforceable primary standards for 21 organic contaminants, 3 inorganic contaminants (arsenic, lead, thallium), 2 disinfection byproduct (DBP) contaminants (bromate, chlorite) and 4 radionuclides (combined radium-226 and -228, gross alpha particles, beta particles and uranium) that have MCLs above their MCLG. Contaminants such as arsenic, lead and many organics have MCLGs of zero; for these contaminants, the MCL will always be higher than the MCLG. Total trihalomethanes (TTHM) and the sum of five haloacetic acids (HAA5) are regulated DBP class sums; while TTHM and HAA5 do not have MCLGs, some of the individual DBP species that comprise these class sums have MCLGs below the TTHM or HAA5 MCL. NPDWRs also include microorganism contaminants (*Cryptosporidium*, *Giardia lamblia*, *Legionella*, total coliforms, including fecal coliform and *E. coli*, and viruses) with MCLGs of zero.

The USEPA's Third Six Year Review (SYR3) assessed 76 NPDWRs promulgated prior to August 2008 and 12 other NPDWRs (e.g. lead, copper, trichloroethylene (TCE) and tetrachloroethylene (PCE)) that were in review at the time of the SYR3. Publicly available data for contaminants under these regulations were collected from 2006 through 2011 for drinking water systems of all sizes. Prior to this study, the SYR3 data sets represented the most recent and comprehensive national occurrence data for most regulated contaminants, including organic contaminants, inorganic contaminants (e.g. arsenic), some DBPs (e.g. bromate, chlorite), radionuclides, and microorganisms.

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 public water systems (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. In summary, drinking water quality data were received from a total of 46 states.

Data were also collected from the USEPA's Fourth Unregulated Contaminant Monitoring Rule (UCMR4) for project analytes, including HAA5, the sum of six brominated haloacetic acids (HAA6Br), and the sum of nine haloacetic acids (HAA9). Data from the USEPA's UCMR4 for specific DBP analytes, including HAA5, HAA6Br, and HAA9, for collection dates during 2018 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. 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 states 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, the 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. Data records with a unit of measure that was not the standard unit of measure or one of the acceptable units of measure were flagged as having non-sensical units and were not included in the final version of the data set that was uploaded to the project database. An exemption to this was made for non-detect data and microbial data. Most data sets did not assign a unit of measure to non-detect data. Non-detect data were assigned the standard unit of measure and a concentration of 0.

For all analytes except microbial analytes, a range of plausible detected concentrations were assigned. Statistical analyses of data outliers, a review of USEPA's 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. For analytes with a USEPA federal-level MCL, an upper threshold of 15 times the MCL was used, where any data record with a concentration greater than 15 times the MCL was assigned an out of range flag. The lower thresholds were determined on an individual analyte basis.

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 non assigned an out of range flag were included in the project database.

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 130,000 drinking water systems and approximately 47.4 million data records.

Data Analysis and Results

The data analysis effort focused on meeting the project objective of assessing the concentration and frequency of, and population affected by, the occurrence of drinking water contaminants above the MCLG but below the MCL or action level with a focus on arsenic, lead, and DBPs, specifically TTHM and HAA5.

Arsenic

Arsenic has an MCLG of zero and an MCL of 10 µg/L. For the 64,694 total drinking water systems with available arsenic occurrence data for 2009 – 2018, the national median concentration was non-detect and the national 95th percentile was equal to 17 µg/L. A total of 10,469 or 28% of community water systems (CWSs), serving a total population of approximately 64 million, had a system-wide median arsenic concentration above the detection level, and therefore above the MCLG of zero, and equal to or below the MCL of 10 µg/L. There were 14,518 or 39% of CWSs, serving a total population of approximately 124 million, with a system-wide 95th percentile arsenic concentration above the MCLG and detection level and equal to or below the MCL of 10 µg/L. A total of 6,174 or 31% of non-transient non-community water systems (NTNCWSs) and transient non-community water systems (TNCWSs), serving a combined population of approximately 1.7 million, had a system-wide median arsenic concentration above the MCLG of zero and the detection level but below or equal to the MCL of 10 µg/L. There were 7,324 or 36% of NTNCWSs and TNCWSs, serving a total population of approximately 2.1 million that had a 95th percentile arsenic concentration above the MCLG and below or equal to the MCL.

Comparing annual 95th percentile arsenic concentrations by system size categories showed that high levels of arsenic tend to be more problematic for smaller systems compared with larger systems. Very small systems, serving populations less than 500, had a 95th percentile concentration equal to 22 µg/L, more than double the MCL of 10 µg/L, in both 2009 and 2014. Very large systems, serving populations greater than 100,000, had the lowest annual 95th percentile concentrations, just above or equal to the MCL in most years from 2009 through 2018. Comparing system types showed that that high levels of arsenic are more of an issue for NTNCWSs, which are regulated by the Arsenic Rule, and in TNCWSs, which are not regulated by the Arsenic Rule, than for CWS. Among the different source water types, groundwater systems have the highest 95th percentile arsenic levels in the period from 2009 through 2018, followed by surface water systems and then groundwater under the direct influence of surface water systems.

Lead

Lead also has an MCLG of zero and has an action level of 15 µg/L. For the 62,504 drinking water systems with available lead occurrence data, the national median concentration for 2009 – 2018 was non-detect and the national 95th percentile was 7.8 µg/L. A total of 10,888 or 28% of CWSs, serving a total population of approximately 57 million, had a system-wide median lead concentration above the detection level and the MCLG of zero and below or equal to the action level of 15 µg/L, and there were 28,725 or 74% of CWSs, serving a total population of approximately 164 million, with a system-wide 95th percentile lead concentration above the MCLG and below or equal to the action level. There were a total of 5,788 or 33% of NTNCWSs and TNCWSs, serving a total population of approximately 1.7 million, with a system-wide median lead concentration above the detection level and the MCLG of zero and below or equal to the action level of 15 µg/L, and there were 10,653 or 61% of NTNCWSs and TNCWSs, serving a total population of approximately 3.5 million, with a system-wide 95th percentile lead concentration above the MCLG and below or equal to the action level. The LCR requires monitoring at taps in homes/buildings that are at a high risk of lead or copper contamination, and it is assumed that available lead occurrence data are entirely or predominantly LCR monitoring data. If every tap at every home/building were sampled, the percentiles for national lead occurrence would be expected to be lower.

The annual 95th percentile lead concentrations in 2009 through 2015 were relatively similar across different system size categories, while for 2016 through 2018, the annual 95th percentile concentrations for very large systems were higher than for systems of other sizes. Further investigation is needed to determine the reason for the increase in annual 95th percentile lead concentrations for very large systems after 2015. The annual 95th percentile concentration for different system types and primary source water types are also relatively similar, although CWSs have consistently had the lowest annual 95th percentile concentration of the three system type categories, and in recent years (2016 – 2018), groundwater systems have had the lowest annual 95th percentile concentrations compared with other two primary source water types. Note that TNCWSs are not required to conduct monitoring at customer taps under the LCR, while CWSs and NCNTWSs are required to do so.

TTHM and HAA5

TTHM and HAA5 are class sums of DBPs and do not have MCLGs. TTHM has an MCL of 80 µg/L and HAA5 has an MCL of 60 µg/L. For the 54,533 drinking water systems with TTHM occurrence data, the national median concentration for 2009 – 2018 was 24 µg/L, and the 95th percentile was 83.6 µg/L, exceeding the MCL of 80 µg/L. For the 46,772 systems with available HAA5 occurrence data, the national median concentration for 2009 – 2018 was 14.7 µg/L and the 95th percentile was 53 µg/L.

A total of 38,599 CWSs, or 99% of CWSs included in this analysis, serving a total population of approximately 271.8 million, had a system-wide median TTHM concentration below or equal to the TTHM MCL of 80 µg/L for 2009 through 2018, and 34,303 or 88% of CWSs, serving a total population of approximately 225.4 million, had a system-wide 95th percentile TTHM concentration below or equal to the MCL. A total of 10,990 NTNCWSs and TNCWSs, or 99.7% of NTNCWSs and TNCWSs included in this analysis, serving a total population of 4.6 million, had a system-wide median concentration below the TTHM MCL of 80 µg/L for 2009 through 2018, and 10,480 or 95% of NTNCWSs and TNCWSs, serving a total population of approximately 4.2 million, had a system-wide 95th percentile TTHM concentration below the TTHM MCL.

A total of 35,675 CWSs, or 99.8% of CWSs included in this analysis, serving a total population of approximately 282 million, had a system-wide median HAA5 concentration below or equal to the HAA5 MCL of 60 µg/L for 2009 through 2018. There were 33,183 CWSs, or 93% of CWSs, serving a total population of approximately 260 million, with a system-wide 95th percentile HAA5 concentration below or equal to the HAA5 MCL. A total of 7,074 or 99.8% of NTNCWSs and TNCWSs included in this analysis, serving a total population of approximately 3.8 million, had a system-wide median HAA5 concentration below or equal to the HAA5 MCL of 60 µg/L for 2009 through 2018. There were 6,831 or 96% of NTNCWSs and TNCWSs, serving a total population of approximately 3.6 million, with a system-wide 95th percentile HAA5 concentration below or equal to the HAA5 MCL.

The annual 95th percentile TTHM and HAA5 concentrations were higher for smaller systems as compared with larger systems. One exception is for very small systems, which had lower annual 95th percentile DBP concentrations as compared with small systems and, in some years, medium systems. This may be due to most very small systems being solely groundwater systems with relatively small distribution systems. Compared with surface water, groundwater tends to have lower levels of total organic carbon (TOC), which corresponds to a lower TTHM and HAA5 formation potential. DBPs continue to form throughout the distribution system as long as a chlorine residual is present, so a smaller distribution system can reduce DBP formation. The annual 95th percentile TTHM and HAA5 concentration tends to be the highest for CWSs. For TTHM, TNCWSs consistently had the lowest annual 95th percentile concentration, but for HAA5, NTNCWSs have had the lowest annual 95th percentile concentration since 2012. Note that the Stage 1 and Stage 2 DBPRs apply to all CWSs and NTNCWSs that add/deliver a primary or residual disinfectant, and only TNCWSs that use chlorine dioxide. Surface water systems have the highest annual 95th percentile TTHM and HAA5 concentration, while groundwater systems and groundwater under the influence of surface water systems have similar 95th percentile DBP concentrations. This result is consistent with the fact that surface waters generally have higher levels of TOC and therefore have a higher formation potential for TTHM and HAA5.

Gaps and Opportunities for Additional Research

The study described in this report focused on data collection for drinking water contaminants with an MCL greater than the corresponding MCLG from state regulatory agencies and from the USEPA's UCMR4 data set, data QA/QC, the development of a database to house these data, and a preliminary data review and analysis to summarize national occurrence and populations affected. The effort leaves research gaps, which present important opportunities for future investigation to better understand the national occurrence of these contaminants in drinking water provided by PWSs.

The QA/QC process did not include a sample location analysis, which would allow for characterizing data records as either "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 customer's taps. For regulated contaminants with an MCLG below the MCL or action level, compliance samples collected by the utility and submitted to the state regulatory agencies should be in the treated or 'finished' water at either the entry point to the distribution system or at distribution system sample locations (e.g. disinfection byproducts) based on compliance monitoring requirements. It is expected that the far majority of data collected as part of this effort were results for finished water samples, either taken at the entry point to the distribution system, at sample points throughout the distribution system (i.e. DBPs, microbial data), or at customer taps (i.e. lead). Further investigation should be conducted to determine the extent of data collected in raw versus finished water. The outcome of such an investigation would provide a better understanding of how representative the results provided in this report are of the water quality delivered to consumers. Once the sample location analysis is completed, a more detailed study of parameters of interest would provide a greater understanding of contaminant occurrence in drinking water delivered to consumers across the US. For example, an investigation of lead data could be conducted to determine if the data allow for calculating 90th percentiles that correspond with compliance calculations and resulting LCR violations.

Co-occurrence analyses, using contaminants from this study and analytes from the Water Quality Research Foundation (WQRF) Aesthetics Level Occurrence Study, such as chlorine and chloramine residual and pH, could provide meaningful insight to national drinking water quality. For example, a detailed analysis of the co-occurrence of DBPs and chlorine/chloramine residual could be conducted. A comparison of DBP occurrence among systems utilizing chloramines compared to those using free chlorine for secondary disinfection could also be conducted, as well as studying DBP occurrence over time for systems that have changed from utilizing free chlorine to chloramines. An investigation of lead and/or arsenic occurrence with corresponding pH data could be implemented as well.

Additionally, the data collection effort collected data for over 1,000 analytes, although the QA/QC, database, and data analysis for this project included the contaminants presented in this report and the aesthetic analytes presented in the WQRF Aesthetics Level Occurrence Study. These additional analytes require data formatting and QA/QC before they can be incorporated into the national database and used for data analyses. A future effort to incorporate those analytes, which include contaminants of national interest, such as nitrate, per- and polyfluoroalkyl substances (PFAS), and many more, would be an important contribution to better understanding national drinking water quality.