Environ Health Perspect

DOI: 10.1289/EHP9985

Supplemental Material

Estimating National Exposures and Potential Bladder Cancer Cases Associated with Chlorination DBPs in U.S. Drinking Water

Richard J. Weisman, Austin Heinrich, Frank Letkiewicz, Michael Messner, Kirsten Studer, Lili Wang, and Stig Regli

Table of Contents

Additional Information About USEPA DBP Regulations.

The interim Total Trihalomethanes (TTHM) Rule established a maximum contaminant level (MCL) of 100 µg/L for TTHM and applied to PWSs that served more than 10,000 people.[1] In this paper, we refer to TTHM as THM4 – i.e., the sum concentration of four trihalomethanes: chloroform, bromodichloromethane, chlorodibromomethane, and bromoform.

The Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules (D/DBPRs) applied to PWSs of all sizes, including, for the first time, those serving fewer than 10,000 people.[2,3] It included an MCL of 80 µg/L for THM4 and an MCL of 60 µg/L for HAA5 – i.e., the sum concentration of the following five haloacetic acids: monochloroacetic, dichloroacetic, trichloroacetic, monobromoacetic and dibromoacetic acids.

The Stage 1 D/DBPR MCLs for THM4 and HAA5 were established as a running annual average of their concentrations combined from all monitoring sites in the distribution system (40 CFR 141.131(b)(2) indicates using zero for analytical results that are less than the Minimum Reporting Level [MRL] concentration). It also included a requirement that PWSs using surface water as a source and providing conventional treatment (coagulation, sedimentation, and filtration) control total organic carbon, a recognized precursor for DBP formation, in their water.

The Stage 2 D/DBPR was based on the use of a locational running annual average (LRAA) for THM4 and HAA5 (i.e., MCL values that must be met at each individual monitoring location in the distribution system). Moreover, the Stage 2 D/DBPR monitoring locations were identified based on the use of an Initial Distribution System Evaluation (IDSE) that was designed to include locations with higher occurrence levels than under the Stage 1 D/DBPR.

In its Economic Analysis (EA) for the Stage 2 D/DBPR, USEPA described its methodology for estimating the compliance activities (e.g., treatment changes) and resulting reductions in DBP concentrations that would result from implementation of that rule.[4] The Stage 1 and Stage 2 regulations were developed to build on previous efforts to reduce exposure to DBPs in drinking water. In the development of the Stage 2 regulations, USEPA estimated that the reductions in exposures would not be as substantial as were achieved by the Stage 1 regulations. As USEPA noted, the reductions in THM4 concentrations and associated bladder cancer cases for compliance activities would not occur immediately due to the cessation lag of the effect of DBP regulations on both existing and future populations and would take several decades to be achieved in full.

USEPA has historically considered microbial protection during the development of the DBP regulations. Given that, USEPA has promulgated microbial regulations simultaneously with DBP regulations. For example, the Interim Enhanced Surface Water Treatment Rule (IESWTR)[5] was promulgated with the Stage 1 D/DBPR, and the Long-Term 2 Enhanced Surface Water Treatment Rule[6] was promulgated with the Stage 2 D/DBPR. Under the IESWTR, procedures were included to help states evaluate potential changes to disinfection practices to not

compromise microbial protection. Those procedures, shown in 40 CFR 141.72, require PWSs to conduct disinfection profiling and benchmarking, which states can use when determining whether a proposed change in disinfection practice is acceptable. The procedures were developed to protect from *Giardia* and viruses.[7]

National Cancer Institute Bladder Cancer Data Description.

The National Cancer Institute (NCI) 2017 estimate of 79,030 for the total number of annual bladder cancer cases from all causes[8] differs from the estimate of 56,506 for 2004 used in the Stage 2 EA.[9] This corresponds to an average annual increase in new bladder cancer cases of approximately 2.6% over this period, even though NCI also notes that the rate of new bladder cancer cases on a population basis has declined by 0.9% per year for the last 10 years. Based on data from the U.S. Census Bureau, between the years 2004 and 2016 (the latest year for which data were available prior to completing this study), the overall U.S. population grew at only an average annual rate of approximately 0.8%.[10,11] However, the population for those aged 55 and older has risen at an annual average rate of approximately 2.7%. Approximately 92% of new bladder cancer cases are in this population group.[12] Therefore, this suggests that the difference in total bladder cancer cases from all causes between the two analyses is likely due to the aging population where most of the new bladder cancer cases occur notwithstanding the slight decrease noted in the rate of new cases.

Additional Information About Post-Stage 2 THM4 Concentration Estimates.

Seidel et al. obtained DBP occurrence data from 395 large systems that each serve more than 100,000 people, across forty-four states, including both surface water and groundwater systems.[13] Their study, which was conducted to assess the impacts from the Stage 2 D/DBPR, showed that the average concentration for THM4 for all sample measurements provided by those systems was 30.5 µg/L from April 2012 to February 2015.[14] The mean of 30.5 µg/L value provided by Seidel et al. was about 5% higher than the 29.1 µg/L estimated from our data (using results for both surface water and groundwater systems). While the data from Seidel et al. showed little change in the median and central tendency (25-75th percentiles) of the measured THM4 concentrations over the past decade, it did show a reduction in the 95th percentile concentrations. Despite differences in methodologies, their findings about reductions in the upper end of the distribution are consistent with our findings about systems with concentrations within the >40-80 µg/L range. Seidel et al. noted several factors that affect the ability of their study to quantify the impacts of the Stage 2 D/DBPR, including factors related to the implementation of the Initial Distribution System Evaluation, the date that systems made treatment modifications to comply with the Stage 2 D/DBPR and other regulations, and required changes in sample collection locations between the Stage 1 and Stage 2 D/DBPR.

Many of the large systems, such as the ones studied by Seidel et al., may have post-Stage 2 sampling locations different from the ones in SYR3. For example, the post-Stage 2 data in Seidel et al. are based on locational RAA sampling locations which are intended to represent values in specified distribution system locations that are generally thought to be higher than average, rather than distribution system-wide average values. Under the Stage 2 D/DBPR, some PWSs provided additional controls to reduce their DBP concentrations, however, USEPA estimated that these controls would not be as extensive as the ones used for implementing the Stage 1 D/DBPR. In addition, the mean values estimated by Seidel et al. are sample-level, rather than system-level, did not adjust system mean concentrations greater than 80 µg/L, and did not distinguish between surface and groundwater systems as was done for this study. While the data summarized by Seidel et al., which were not used in our study, are more recent than that collected under the SYR3, we suggest that the SYR3 data provide a more comprehensive view of mean THM4 exposure for the U.S. population. The SYR3 data reflect measured concentrations in small, medium, and large PWS across the U.S. and are based on post-Stage 1 compliance activities which show distribution system average concentrations.

Table S1: Additional information about number of THM4 records for 2011 in the third Six-Year Review (SYR3) occurrence dataset.[15]

System Size	Source Water Type	Number of Systems	Total Number of Records	Average Number of Records per System
<10,000	GW	13,193	20,400	1.5
	SW	4,824	23,150	4.8
10,000 - 100,000	GW	971	8,944	9.2
	SW	1,471	22,565	15.3
>100,000	GW	58	1,734	29.9
	SW	243	8,954	36.8
Total		20,760	85,747	4.1

Table S2: Relative Difference in ICR Data Between States that Provided SYR3 THM4 Data vs. States that Did Not Provide SYR3 THM4 Data for Evaluation of Representativeness of THM4 Concentrations.

Mean concentration (µg/L)	States included	States not included	Percent difference between states that provided SYR3 THM4 data vs. states that did not provide SYR3 THM4 data
THM4	37.2	39.8	6.8%
HAA5	25.6	26.0	1.3%

Table S2 Note: We used the ICR data set to compare the THM4 concentrations from systems among the eight missing SYR3 states (27 systems) with that of systems from the remaining forty-two states (223 systems).[15,16] There was a 6.8% difference for THM4 and a 1.3% difference for HAA5 between states that provided SYR3 THM4 data and states that did not provide such data.

R Script of Data and Equations for Figure #3 (R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/).

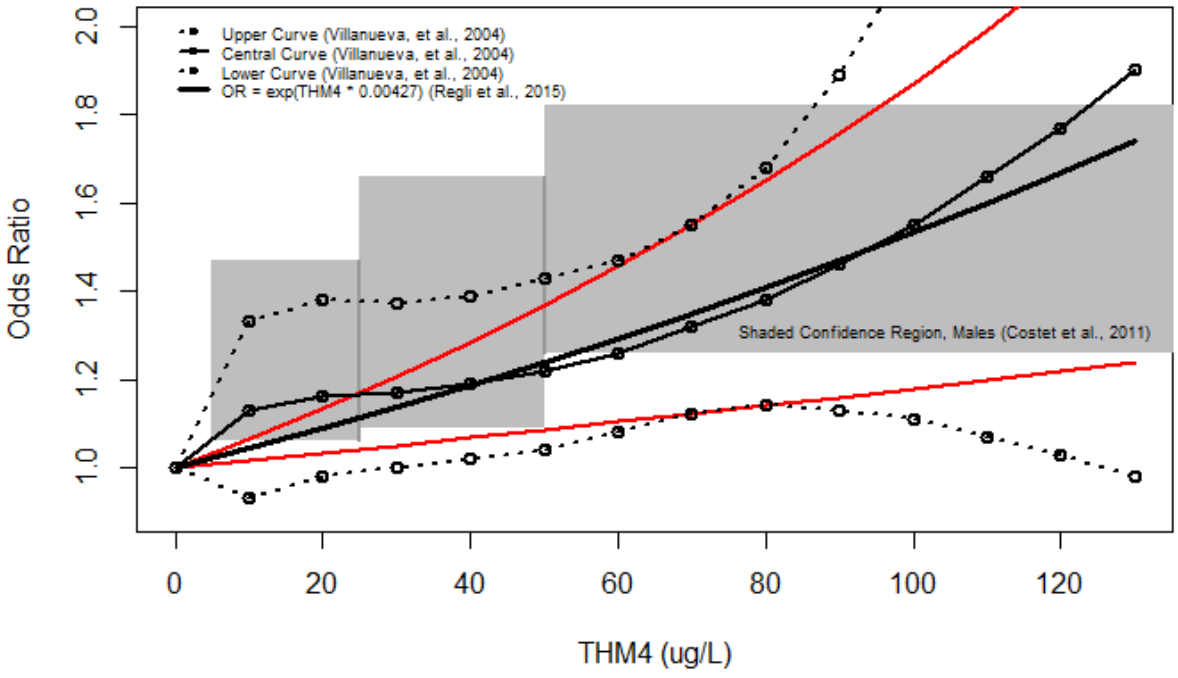
```
#  Table 1 of Regli et al. (2015)
TTHM <- seq(from = 0, to = 130, by = 10)
OR <- c(1.00, 1.13, 1.16, 1.17, 1.19, 1.22, 1.26, 1.32, 1.38,
      1.46, 1.55, 1.66, 1.77, 1.9)
LCL <- c(1.00, 0.93, 0.98, 1.00, 1.02, 1.04, 1.08, 1.12, 1.14,
      1.13, 1.11, 1.07, 1.03, 0.98)
UCL <- c(1.00, 1.33, 1.38, 1.37, 1.39, 1.43, 1.47, 1.55, 1.68,
      1.89, 2.17, 2.55, 3.06, 3.66)

# Odds ratio as function of concentration
Odds <- function(P) P / (1 - P)
b <- 0.00581        # Slope of OR function
ORcentral <- function(C) 1 + b * C
ORcentral(38.05)    # 1.22107
b.2 <- 0.004266733  # was 0.00581 (from fit to unweighted points)
ORcentral.2 <- function(C) exp(C * b.2)  # above: 1 + b.2 * C

# Create Figure 3. Estimates from Costet's Table 4 are entered in the polygon
# commands, below. There is one polygon for each of the table's three non-
# referent exposure ranges. (The referent range is 0 to 5 ug/L.)
# Empty plot box, with axes labeled
plot(TTHM, OR, xlab = "THM4 (ug/L)", ylab = "Odds Ratio",
     ylim = c(0.9, 2), xlim = c(0, 130), # log = "y", cex = 0.8,
     type = "o", lwd = 2, col = "white")

# Add polygons from Costet et al., Table 4.
polygon(x = c(5, 25, 25, 5), y = c(1.06, 1.06, 1.47, 1.47),
        col = "gray", border = NA)
polygon(x = c(25, 50, 50, 25), y = c(1.09, 1.09, 1.66, 1.66),
        col = "gray", border = NA)
polygon(x = c(150, 50, 50, 150), y = c(1.26, 1.26, 1.82, 1.82),
        col = "gray", border = NA)

# Add vertical gray lines to make the three rectangles stand out
points(c(25, 25), c(1.06, 1.47), type = "l", col = "gray60", lwd = 2)
points(c(50, 50), c(1.26, 1.66), type = "l", col = "gray60", lwd = 2)

# Label the gray region
text(74, 1.31, "Shaded Confidence Region, Males (Costet et al., 2011)",
     pos = 4, cex = 0.55, # cex = 0.9, if using windows(10, 7)
     col = "black")

# Add Villanueva's three curves
points(TTHM, OR, type = "o", lwd = 2, col = "black")
points(TTHM, LCL, type = "o", lty = 3, lwd = 2, col = "black")
points(TTHM, UCL, type = "o", lty = 3, lwd = 2, col = "black")

# Add Regli et al. curve (using corrected fitted slope)
points(TTHM, ORcentral.2(TTHM),
      type = "l", pch = "+", col = "black", lwd = 3)

# Add legend
```

```
legend(0, 2.03, bty = "n", cex = 0.56, # cex = 0.9, if using windows(10, 7)
       c("Upper Curve (Villanueva, et al., 2004)",
         "Central Curve (Villanueva, et al., 2004)",
         "Lower Curve (Villanueva, et al., 2004)",
         "OR = exp(THM4 * 0.00427) (Regli et al., 2015)"),
       col = c("black", "Black", "black", "black"),
       lty = c(3, 1, 3, 1),
       pch = c(1, 1, 1, NA),
       lwd = c(2, 2, 2, 3))
```

R Script for Deriving the Model Parameters (Limiting Slopes) Based on the Lower- and Upper-95 Percent Confidence Interval Curves in Villanueva et al. (2004)

```
# First, Slope.95 of the function *exp(Slope.95 $\times$ TTHM)* is found
# so that the function # falls at or below each of Villanueva's upper set
# of points. This "slope" (`r round(log(UCL[8]) / # TTHM[8], 5) `) defines
# the upper end of our confidence interval. Note that this function
# passes through the upper point at TTHM = 70.

# Add the new upper curve (based on the 8th ordered pair, TTHM[8] = 70)
Slope.95 <- log(UCL[8]) / TTHM[8]

# Next, Slope.05 of the function *exp(Slope.05 $\times$ TTHM)*
# is found so that the function
# falls at or above each of Villanueva's lower set of points.
# This "slope" (`r round(log(LCL[9]) / TTHM[9], 5) `) defines
# the upper end of our confidence interval.
# Note that this function
# passes through the lower point at TTHM = 80.

# Add the new lower curve (based on the 9th ordered pair, TTHM[9] = 80)
Slope.05 <- log(LCL[9]) / TTHM[9]

# Include the two curves with the plot
plot(TTHM, OR, xlab = "THM4 (ug/L)", ylab = "Odds Ratio",
     ylim = c(0.9, 2), xlim = c(0, 130), # log = "y", cex = 0.8,
     type = "o", lwd = 2, col = "white")

# Add polygons from Costet et al., Table 4.
polygon(x = c(5, 25, 25, 5), y = c(1.06, 1.06, 1.47, 1.47),
        col = "gray", border = NA)
polygon(x = c(25, 50, 50, 25), y = c(1.09, 1.09, 1.66, 1.66),
        col = "gray", border = NA)
polygon(x = c(150, 50, 50, 150), y = c(1.26, 1.26, 1.82, 1.82),
        col = "gray", border = NA)

# Add vertical gray lines to make the three rectangles stand out
points(c(25, 25), c(1.06, 1.47), type = "l", col = "gray60", lwd = 2)
points(c(50, 50), c(1.26, 1.66), type = "l", col = "gray60", lwd = 2)

points(c(TTHM), c(exp(TTHM * Slope.95)), type = "l", col = "red", lwd = 2)
points(c(TTHM), c(exp(TTHM * Slope.05)), type = "l", col = "red", lwd = 2)

# Label the gray region
text(74, 1.31, "Shaded Confidence Region, Males (Costet et al., 2011)",
     pos = 4, cex = 0.55, # cex = 0.9, if using windows(10, 7)
```

```
     col = "black")

# Add Villanueva's three curves
points(TTHM, OR, type = "o", lwd = 2, col = "black")
points(TTHM, LCL, type = "o", lty = 3, lwd = 2, col = "black")
points(TTHM, UCL, type = "o", lty = 3, lwd = 2, col = "black")

# Add our curve (using corrected fitted slope)
points(TTHM, ORcentral.2(TTHM), type = "l",
      pch = "+", col = "black", lwd = 3)

# Add legend
legend(0, 2.03, bty = "n", cex = 0.56, # cex = 0.9, if using windows(10, 7)
       c("Upper Curve (Villanueva, et al., 2004)",
         "Central Curve (Villanueva, et al., 2004)",
         "Lower Curve (Villanueva, et al., 2004)",
         "OR = exp(THM4 * 0.00427) (Regli et al., 2015)"),
       col = c("black", "Black", "black", "black"),
       lty = c(3, 1, 3, 1),
       pch = c(1, 1, 1, NA),
       lwd = c(2, 2, 2, 3))
```



References:

1. USEPA. 1979. National Interim Primary Drinking Water Regulations; Control of Trihalomethanes in Drinking Water; Final Rule. 44 FR 68624. November 29.
2. USEPA. 1998a. National Primary Drinking Water Regulations: Disinfectants and Disinfection Byproducts; Final Rule. 63 FR 69390. December 16.
3. USEPA. 2006a. National Primary Drinking Water Regulations: Stage 2 Disinfectants and Disinfection Byproducts Rule; Final Rule. 71 FR 388. January 4.
4. USEPA. 2005. Economic Analysis for the Final Stage 2 Disinfectants and Disinfection Byproducts Rule, EPA 815-R-05-010. December.
5. USEPA. 1998b. National Primary Drinking Water Regulations: Interim Enhanced Surface Water Treatment; Final Rule. 63 FR 69478. December 16.
6. USEPA. 2006b. National Primary Drinking Water Regulations: Long-Term 2 Enhanced Surface Water Treatment Rule; Final Rule. 71 FR 654. January 5.
7. USEPA. 1999. Disinfection Profiling and Benchmarking Guidance Manual. EPA 815-R-99-013. August.
8. NCI. 2018. SEER Cancer Stat Facts: Bladder Cancer. Bethesda, MD. https://seer.cancer.gov/statfacts/html/urinb.html. Downloaded January 2.
9. USEPA. 2005. Economic Analysis for the Final Stage 2 Disinfectants and Disinfection Byproducts Rule, EPA 815-R-05-010. December.
10. U.S. Census Bureau. 2017a. State Intercensal Dataset 2000:2010. Intercensal Estimates of the Resident Population by Single Year of Age and Sex for States and the United States: April 1, 2000 to July 1, 2010. Available at: https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html.
11. U.S. Census Bureau. 2017b. Annual estimates of the resident population by single year of age and sex for the United States: April 1, 2010 to July 1, 2016 (NC-EST2016-AGESEX-RES). Available at: https://www.census.gov/data/datasets/2016/demo/popest/nation-detail.html.
12. NCI. 2018. SEER Cancer Stat Facts: Bladder Cancer. Bethesda, MD. https://seer.cancer.gov/statfacts/html/urinb.html. Downloaded January 2.
13. Seidel, C.J., Samson, C.C., Bartrand, T., Ergul, A., Summers, R. S. 2017. "Disinfection Byproduct Occurrence at Large Water Systems After Stage 2 DBPR". Journal of the American Water Works Association. Vol 109, No 7. DOI: https://doi.org/10.5942/jawwa.2017.109.0082**.**
14. Samson, C. 2018. Personal communication. March 23.
15. USEPA. 2018. Six-Year Review 3 Compliance Monitoring Data (2006-2011). Available online at: https://www.epa.gov/dwsixyearreview/contaminant-occurrence-and-related-data-six-year-review-drinking-water-standards. Downloaded February 22.
16. USEPA. 2000. ICR Auxiliary 1 Database. EPA-815-C-00-002. April.