Methods

Indicator

E7. Estimated percentage of children ages 0 to 17 years served by community water systems that did not meet all applicable health-based drinking water standards, 1993–2023.

E8. Estimated percentage of children ages 0 to 17 years served by community water systems with violations of drinking water monitoring and reporting requirements, 1993–2023.

Summary

EPA's Safe Drinking Water Information System Federal Version (SDWIS/FED) includes information on populations served and violations of maximum contaminant levels or required treatment techniques by the nation's over 148,000 public water systems[i]. For each calendar year, SDWIS/FED violation data for the nation's community water systems, which serve most people in the United States, were obtained from EPA Office of Ground Water and Drinking Water (OGWDW), listing all violations that occurred during part or all of that year. For Indicator E7, health-based violations were grouped into different types based on the violation code and chemical contaminant code. For Indicator E8, violations of monitoring and reporting requirements were grouped into different types based on the violation code. For each state and type of violation, the fraction of the population served by violating systems was estimated as the total population served by community water systems with one or more violations divided by the total population served by all community water systems. For each state and type of violation, the number of children served by violating systems was estimated as the fraction of the population served by violating systems multiplied by the total population of children ages 17 years and under in the state. Indicators E7 and E8 are the percentages of children served by violating systems in relation to all children served by community water systems, estimated by summing the numbers of children served by violating systems across all states and dividing this total by the national total population of children ages 17 years and under.

Overview of Data Files

The following files are needed to calculate this indicator:

- SDWIS/FED violation data. This file contains the public water system identification code (PWSID), violation code, contaminant code, contaminant type, state (primary state served by the water system), population served, violation begin date, violation end date, return to compliance date, compliance status code, and compliance status description This file has one row for each violation excluding inapplicable violations with code 500. This file only

---

[i] U.S. Environmental Protection Agency. (2024, October 30). *Information about public water systems.* https://www.epa.gov/dwreginfo/information-about-public-water-systems

contains violations by Community Water Systems (as defined by OGWDW).[ii] The population served is the current (first quarter 2024) estimate of the population served by that water system. This file was obtained directly from OGWDW staff.

- SDWIS/FED inventory data. This file contains the total population served by all active community water systems in each state and year. The data for 2023, first quarter 2024, frozen as of April 2024, were used for these analyses. This file was obtained directly from OGWDW staff.

- Census data. This file contains the state and county FIPS codes, year, and children's population. For 1993-1999, we obtained this information from the U.S. Census Bureau files:

> Estimates of the Population of Counties by Age and Sex: 1990-1999, August 30, 2000. The file headers were "(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999." http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html. These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and county across all ages 0 to 17 years and both sexes.

For 2000-2009, we obtained this information from the bridged-race intercensal population files:

> National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

For 2010-2020, we obtained this information from the bridged race Vintage 2020 postcensal population file:

> National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release by the U.S. Census Bureau of the unbridged 33Vintage 2019 postcensal estimates by 5-year age group on June 17, 2021

[ii] Community Water Systems have at least 15 service connections or serve 25 or more of the same population year-round.

Starting from 2021, because bridged-race Vintage postcensal population data are no longer provided by the National Center for Health Statistics, we obtained this information from the U.S. Census Bureau files:

> "County Characteristics Resident Population Estimates," July 1, 2023. The file header is "Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2023." The file description and data link can be found at https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html. The data set is located at https://www2.census.gov/programs-surveys/popest/datasets/2020-2023/counties/asrh/.

The children's populations by year and state, for July 1, were obtained by summing across all counties and across the ages 0 to 17 inclusive.

Drinking Water Standards Violation Data

Drinking water standards violation data from the EPA Safe Drinking Water Information System Federal Version (SDWIS/FED) were obtained directly from OGWDW[iii]. The file was extracted from SDWIS/FED in January 2025. This file contains the public water system identification code (PWSID), violation code, contaminant code, state (primary state served by the water system), population served, violation begin date, violation end date, return to compliance date, compliance status code, and compliance status description. This file has one row for each violation excluding inapplicable violations with code 500. This file only contains violations by Community Water Systems. The data included both active and inactive PWS to account for potential historical violations of currently inactive systems in the previous years when those systems were active. The population served is the current (first quarter 2024) estimate of the population served by that water system. Violations by systems not assigned to one of the 50 U.S. States or Washington DC were excluded. Therefore, we did not include violations in Puerto Rico and other U.S. territories that are not states. We also excluded violations on tribal lands which are assigned to the EPA regions. Finally, we excluded a single SWTR TT for PWSID CA1910067 determined to be invalid by OGWDW staff.

Each violation was assigned an effective end date based on the compliance status description, violation end date, and return to compliance date, as follows:

- Compliance status description = "Known." Return to compliance date is missing. Effective end date = violation end date.
- Compliance status description = "Inactive." Return to compliance date is the date the system became inactive. If the return to compliance date is before the violation begin date, then exclude that violation. Otherwise effective end date = earliest of violation end date and return to compliance date.

[iii]Chandler Klawitter, 202-564-1954

- Compliance status description = "Returned to compliance." Effective end date = earliest of violation end date and return to compliance date.
- Compliance status description = "Open." Return to compliance date and violation end date are both missing. Effective end date = 31 December, 2023.

Each violation was assigned to each and every calendar year that intersects the period from the violation begin date to the effective end date. Thus the violation is assigned to each calendar year such that the violation begin date is on or before December 31 and the effective end date is on or after January 1.

For Indicator E7, health-based violations were grouped into the following violation types using the violation and contaminant codes:

- All health-based: Violation codes 41, 42, 45, and 48 with contaminant code 700 (Ground water rule), 40 (Filter backwash recycling rule), or any of the following violations
  - Lead and copper: Violation codes 57, 58, 59, 63, 64, 65
  - Total coliforms: Violation codes 1 and 2 with contaminant code 3000 (Coliform pre-TCR) only. Total Coliform Rule: Violation codes 21 and 22 with all contaminant codes. Revised Total Coliforms Rule (April 2016): Violation codes 1A, 2A, 2B, 2C and 2D with contaminant code 8000 only*
  - Chemical and radionuclide: Violation codes 1, 2 and 7 with all applicable contaminant codes**
  - Surface water treatment: Violation codes 33, 37, 41-45, and 47 with all contaminant codes except 700
  - Nitrate/nitrite: Violation codes 1 and 2 with contaminant codes 1038, 1040, and 1041 only
  - Disinfectants and disinfection byproducts: Violation codes 11, 12, 13, and 46 with all contaminant codes. Violation code 2 with contaminant codes 1009, 1011, 2456, and 2950 only

* In 2016, the Total Coliform Rule was replaced by the Revised Total Coliform Rule which began implementation in April 2016.
** The applicable contaminant codes for Chemical and radionuclide are as follows:

Volatile organic chemicals: 2265, 2257, 2378, 2380, 2955, 2964, 2968, 2969, 2976, 2977, 2979, 2980, 2981, 2982, 2983, 2984, 2985, 2987, 2989, 2990, 2991, 2992, 2996
Synthetic organic contaminants: 2005, 2010, 2015, 2020, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2039, 2040, 2041, 2042, 2046, 2050, 2051, 2063, 2065, 2067, 2105, 2110, 2274, 2306, 2326, 2383, 2931, 2946, 2959
Inorganic chemicals: 1005, 1010, 1015, 1020, 1024, 1035, 1036, 1045, 1074, 1075, 1085, 1094
Radiological contaminants: 4000, 4006, 4010, 4100, 4101, 4102, 4174, 4264

For Indicator E8, monitoring and reporting violations were grouped into the following violation types using the violation codes:

- Any violation: Violation codes 19, 34 with all contaminant codes and 31 with contaminant code 700 (Ground water rule), 39 (Filter backwash recycling rule), or any of the following violations
  - Lead and copper: Violation codes 51, 52, 53, 56, and 66
  - Total coliforms: Total Coliform Rule: Violation codes 23, 24, 25, and 26. Revised Total Coliforms Rule (April 2016): Violation codes 3A, 3B, 3C, 3D, 4A, 4B, 4C, 4D, 4E and 4F with contaminant code 8000 only***
  - Chemical and radionuclide: Violation codes 3 and 4
  - Surface water treatment: Violation codes 29, 32, 36 and 38 with all contaminant codes. Violation code 31 with contaminant codes 200 and 800 only
  - Disinfectants and disinfection byproducts: Violation codes 27, 30, and 35

*** In 2016, the Total Coliform Rule was replaced by the Revised Total Coliform Rule which began implementation in April 2016.

In a calendar year, a PWSID may have one or more violations of a given violation type. Only the first such violation is selected. For each violation type, state, and calendar year, the total population served by all violating PWSIDs was calculated by summing across the PWSIDs, counting each violating PWSID once only.

For each state, the total population served by community water systems was obtained from the SDWIS/FED inventory files. OGWDW staff provided this summary file. The data for 2023, first quarter 2024, frozen as of April 2024, were used for these analyses. Only the 50 U.S. States or Washington DC, identified by their postal code abbreviations, were included.

Census Data

We obtained children's populations by state for each year from 1993-2019.

For 1993-1999, the source was U.S. Census Bureau files:

> "Estimates of the Population of Counties by Age and Sex: 1990-1999," August 30, 2000. The file headers are "(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999." The file headers were "(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999."
> http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html.

These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and state across all counties in the state, all ages 17 years and under, and both sexes.

For 2000-2009, we used the bridged-race intercensal population files obtained from the CDC website:

> National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: https://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

The bridged-race intercensal population files contain estimates of the resident population of the United States as of July 1, 2000, July 1, 2001, July 1, 2002, July 1, 2003, July 1, 2004, July 1, 2005, July 1, 2006, July 1, 2007, July 1, 2008, and July 1, 2009 by county, single-year of age (0, 1, 2,..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. There is one SAS file for the years 2000-2004 and another SAS file for the years 2005-2009. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the variables: state, county, age, racesex, hisp, and pop2000 to pop2009. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX gives the population as of July 1 of the calendar year XXXX for a given state, county, age, racesex combination, and ethnicity. The state children's populations for each year 2000-2009 were obtained by summing the variable popXXXX over all counties in a state, all ages 17 years and under, all values of "racesex" and all values of "hisp."

For 2010-2020, we used the bridged race Vintage 2020 postcensal population files obtained from the CDC website:

> National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from https://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release by the U.S. Census Bureau of the unbridged Vintage 2019 postcensal estimates by 5-year age group on June 17, 2021.

The bridged race 2010 to 2020 population files contain estimates of the resident population of the United States as of July 1, 2010, July 1, 2011, July 1, 2012, July 1, 2013, July 1, 2014, July 1, 2015, July 1, 2016, July 1, 2017, July 1, 2018, July 1, 2019, and July 1, 2020 (the data for April 1, 2010 were not used for these analyses) by county, single-year of age (0, 1, 2,..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the following variables: state, county, age, racesex, hisp, pop2010_jul, pop2011, pop2012, pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, and pop2019. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX or popXXXX_jul gives the population as of July 1, XXXX for a given state, county, age, racesex combination, and ethnicity. The state children's populations for the years 2010 to 2019 were obtained by summing the variables pop2010_jul, pop2011, pop2012, pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, and pop2019 over all counties in a state, all ages 17 years and under, all values of "racesex," and all values of "hisp."

For 2021-2023, the source was U.S. Census Bureau files:

> The file header is "CC-EST2023-ALLDATA-[ST-FIPS]: Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2023" The file is released in July 2023. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html.

This file gives county populations by age groups, sex, unbridged race groups, and Hispanic Origin. We summed these populations by year and state across all counties in the state, all ages 17 years and under, and both sexes.

The unbridged race 2020 to 2023 population file contains estimates of the resident population of the United States as of July 1, 2020, and July 1, 2023. Unbridged race categories (American Indian and Alaska Native alone non-Hispanic, White alone non-Hispanic, Black or African American alone non-Hispanic, Asian alone non-Hispanic, Native Hawaiian and Other Pacific Islander alone non-Hispanic, Two or More Races non-Hispanic, and Hispanic) were used in the analyses of 2023.

We extracted the following variables: state, county, agegrp, and 7 race/ethnicity groups ("NHIA", "NHWA", "NHBA", "NHAA", "NHNA", "NHTOM", and "H") for both males and females. The group "ALL" was the sum of all persons in those 7 race/ethnicity groups. For each state, county, and AGEGP, we summed the male and female populations for each of the following list of race/ethnicity groups which use the unbridged race groupings: "All" = All persons, "AIANNH" = AIAN alone non-Hispanic (NH) , "WNH" = White alone NH, "BNH" = Black or African American alone NH, "ANH" Asian alone NH, "NHOPINH" = Native Hawaiian and Other Pacific Islander alone NH, "TwoNH" = two or more races NH, "H" = Hispanic. The populations of children ages 0 to 17 years were calculated by summing the populations for AGEGRP = 1, 2, and 3, and adding 3/5 of the population for AGEGRP = 4. That is because the AGEGPs are "1" = ages 0 to 4, "2" = ages 5 to 9, "3" = ages 10 to 14, and "4" = ages 15 to 19, assuming the ages 15-19 are equally distributed.

Calculation of Indicator

Indicators E7 and E8 are calculated as follows.

Using the water quality violation data, the PWSIDs violating a given type of violation during each calendar year were listed. For each type of violation and calendar year, only the first violation by a PWSID was selected.

1. For each state, type of violation, and year, the populations served by violating systems were obtained by summing the populations served across the violating PWSIDs:

Violating population (state, violation type, year) = Σ Population served

where this sum is taken over all violating community water systems in the state for the given calendar year.

2. The fractions of the populations served by violating systems were obtained by dividing the violating population in the state for that year by the total population served by all community water systems in the state. The total populations served by community water systems in each state were obtained from the SDWIS/FED inventory data:

State population fraction affected (state, violation type, year) =
  Violating population (state, violation type, year) /
  Total population served (state)

3. The number of affected children in each state was estimated by multiplying the state population fraction affected by the state children's population. This estimate assumes an even geographic distribution over water systems of individuals under the age of 18 in each state for each year. This estimate also assumes that all children are served by community water systems; see below for a discussion of this assumption. This estimate also makes no adjustment for children in one state that are served by water systems that primarily serve another state:

State children affected (state, violation type, year) =
  State population fraction affected (state, violation type, year) ×
  Children 0-17 (state, year)

4. The national number of children served by violating systems was obtained by summing the state numbers of children served by violating systems:

National children affected (violation type, year) =
Σ State children affected (state, violation type, year)

5. The percentage of children served by violating water systems was obtained by dividing the national number of children served by violating water systems by the total number of children in the 50 U.S. states and Washington DC:

Percentage children affected (violation type, year) =
[National children affected (violation type, year) /

Σ Children 0-17 (state, year)] × 100%

where the sum is across all states.

In Step 3, an assumption is made that all children ages 0 to 17 years are served by community water systems. In fact, some of the U.S. population, and hence some children ages 0 to 17 years, are not served by community water systems, so this assumption is unrealistic. To address this issue, suppose instead that a fraction g of children are served by community water systems, where g is approximately the same for every state, but could vary by year. In Step 3, the number of children affected in a given state is given by:

> State children affected (state, violation type, year) =
>> State population fraction affected (state, violation type, year) ×
>> Children 0-17 (state, year) × g(year),

since there are Children 0-17 (state, year) × g(year) children ages 0 to 17 years that are served by community water systems in that state and year, Step 4 is unchanged. In Step 5, the percentage of children served by violating systems in relation to the children served by community water systems is obtained by dividing the national number of children served by violating water systems by the total number of children served by community water systems:

> Percentage children affected (violation type, year) =
> [National children affected (violation type, year) /
> Σ Children 0-17 (state, year) × g(year)] × 100%.

Because the same fraction g(year) appears in both the numerator and denominator, it cancels out. Therefore, even though not all children are served by community water systems, Indicators E7 and E8 estimate the percentage of children served by violating systems in relation to the children served by community water systems.

Questions and Comments

Questions regarding these methods, and suggestions to improve the description of the methods, are welcome. Please use the "Contact Us" link at the bottom of any page in the America's Children and the Environment website.