Chemical Exposure Action Map Methods

Mary Collins, State University of New York (SUNY-ESF)
Jeremy Proville, Environmental Defense Fund (EDF)
Paige Varner, Environmental Defense Fund (EDF)
Last Revised: May 29, 2026

For more information and associated code, please visit our Github: https://github.com/edf-org/TSCA-Map

All steps except for "Processing for Display in EDF's Chemical Exposure Action Map" were done via the Areal Apportionment package. More information on this package can be found here: https://github.com/SarahValencia/ArealApportionment.

Introduction

The goal of this project was to assess the potential impact of facilities releasing high-priority toxic chemicals undergoing assessment and regulation under the Toxic Substances Control Act (TSCA) as well as a group of toxic “forever chemicals,” per- and polyfluoroalkyl substances (PFAS), and urge EPA to assess these chemical exposures and risks cumulatively. This was accomplished by intersecting publicly available demographic data from the American Community Survey (ACS) with publicly available Toxics Release Inventory (TRI) data, which provides the locations of facilities that are required to report toxic chemical releases. From these two data sets we calculated the demographics found within 10 kilometers (km) of each TRI facility that released chemicals between 2018 and 2022 of interest to the Environmental Defense Fund (EDF). This document describes the methods used in detail.

Obtaining Census Data

We obtained census data from the 2020 ACS at the block group level for the continental US in 2020 using the getACS function in the tidyCensus package in R. This function allows users to specify the year, spatial scale, and demographic variables that are accessed. Table 1 contains the codes and descriptions of the demographic variables pulled. Data for these variables were obtained at two different spatial scales, including block group and county. County level data was obtained in order to provide a reference point for comparing the demographics within a given distance of a TRI reporting facility with that of the county that the TRI facility is located in.

In some cases the raw data was summed to make more informative categories, including females of reproductive age (15-49), people aged 65+, children under 5, and people without a high school diploma. All population variables were divided by the total population in order to obtain a percentage. We also pulled economic variables including the median income and median housing price in each block group and county, as well as the number of owner occupied, vacant, and total housing units.

Risk-Screening Environmental Indicators Data

The Environmental Protection Agency's Risk-Screening Environmental Indicators (RSEI) database incorporates information tracked by the EPA via the TRI program. It contains data on the releases of federally managed toxic chemicals that may cause harm to human health or the environment. The RSEI database combines the TRI information on the amount of toxic chemicals released or transferred from facilities as well as additional factors such as the chemical's fate and transport through the environment, each chemical's relative toxicity, and potential human exposure to model the risk associated with different chemical releases. We downloaded version 23.11 of the public release data for this project. Additional information about the data contained in each table of the RSEI database can be found here: https://www.epa.gov/rsei/rsei-data-dictionary-aggregated-grid-cell-geographic-microdata.

Identifying TRI Facilities Releasing Chemicals of Interest

The TRI program currently tracks 787 chemicals across 33 chemical categories. To narrow the scope of the project, EDF provided a list of "Chemicals of Interest" which are undergoing regulation under TSCA (https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/ongoing-and-completed-chemical-risk-evaluations-under). We cross-referenced this list with the data provided in the Chems table (found in the file chemical_data_rsei_v2310.csv on the RSEI website), which provides information on all of the chemicals tracked in the RSEI database, and found data available for 28 of the chemicals on the list. We also included data on a group of 180 per- and polyfluoroalkyl substances (PFAS) releases as this group of chemicals causes the same health harms as many of the 28 priority chemicals undergoing assessment and regulation under TSCA.

Next, our goal was to find all reported releases for those 28 chemicals plus 180 PFAS between 2018 and 2022 and sum them to get a yearly amount for each chemical for each TRI facility. Because the RSEI data is stored in a number of different tables, it was necessary to identify the crosswalks between tables in order merge the necessary information into a single output file. This was achieved in a series of "join" steps, in which the desired columns were selected from each RSEI table to create a new table. This abbreviated table was then merged with others.

A subset of columns was selected to create a Facility table that included location information. Similarly, a subset of columns with desired information was selected to create a new "Submission" table. The "Facility" and "Submission" tables were then merged in order to link information on each TRI facility such as location, NAICS code, etc, to each submitted TRI report, which contains info on the year, chemical, and facility for each reported release (there are multiple releases reported per submission number). Because data on the actual poundage of chemicals released is found in 'release_data_rsei_v2311.csv', the columns for Submission Number, Release Number, Media, and Pounds Released were selected from this table to create a new "Release" table. The Release table was then merged with the info from the Facility and Submission tables. A PFAS identifier was also carried over to the table to identify which chemicals fall under the group 'PFAS'.

Areal Apportionment of Demographics near TRI Facilities

Spatial operations were conducted using the sf package in R, and all data sets were converted to spatial objects using the "st_as_sf" function.

TRI facility location data is provided in units of degrees in the North American Datum of 1983 (NAD83; https://epsg.io/4269) coordinate system, which is the system used by the federal government. For spatial analyses involving distances, such as determining the boundaries of concentric circles with radii specified in kilometers, it is necessary to create a projected copy of the data in which latitude and longitude are converted to metric distances.

For more information, please contact Paige Varner, Healthy Communities Scientist, pvarner@edf.org

Environmental Defense Fund
257 Park Avenue South
New York, NY 10010

A vital Earth. For everyone.
EDF.org
T 212 505 2100

To do this, TRI facilities locations were pulled from the filtered RSEI data. Any entries with NAs in the "Latitude" and "Longitude" columns were removed, as those were located outside the continental US. Facility locations were then transformed using the “st_transform” function in the sf package in R to the World Geodetic System (WGS84; https://epsg.io/3857), which a commonly used metric projection system in meters. All three sets of census data were also projected from NAD83 to WGS84.

After converting all data sets to metric projections, we used the “st_buffer” function to create a 10 km buffer around each TRI facility location. Buffers are polygons representing the area within a given distance of a geometric feature. Because we are treating TRI locations as points, the buffer created is a concentric circle with a radius of 1km. Buffers are polygons representing the area within a given distance of a geometric feature: regardless of whether the input is a point, line or polygon, the output is a polygon. The buffer polygons centered on the TRI facilities were overlaid on the census block group polygons using the “st_intersection” function to determine which block groups had area that fell within the buffer. We then used the function “st_area” to determine the amount of area of each block group that fell within the buffer and converted this to a percentage of the total block group area.

To calculate the number of people within the 10 km buffer we assumed that the people counted within each block group in the ACS were evenly distributed across the entire block group. We multiplied the percentage of each block group that fell within the buffer by the number of people in each variable and added up those numbers across all the block groups near each facility to get the total number of people in each demographic that may be subject to releases by a given facility at that distance. While this assumption is obviously a rough approximation and a source of error, we did not have the resources to use cutting edge techniques such as dasymetric mapping with remotely sensed datasets (e.g. Worldpop or Landscan) to reduce this source of uncertainty. The number of people within each demographic variable were converted to percentages, and the county ACS data was used to determine demographic percentages within the county that each facility was found in for comparison purposes.

This produced a data table summarizing the percent of each demography found within 10 km of each TRI facility in the continental US in 2020. This was merged with the RSEI data above so that users could determine, for any given facility-chemical combination between 2018 and 2022, the number of people of different demographics potentially impacted at each distance as well as the associated chemical releases. These data can be found in the file “DemographicImpactsofTRIv2.csv”.

Processing for Display in EDF’s Chemical Exposure Action Map

Data output in the “DemographicImpactsofTRIv2.csv” file were then processed further in the “TSCA mapping analysis.R” script to stage it for display in EDF’s Chemical Exposure Action Map. Each chemical or chemical group of interest was assigned to 1 or multiple health risk outcomes of interest, based on a literature review: cancer, developmental harm, and asthma. In addition to these categorizations, each chemical or group was assigned a weighting factor (as denoted in tab 3 of the “TSCA Fenceline Map Chem Hazard Data Extractionv2.xlsx" spreadsheet). Cancer chemicals were weighted based on the International Agency for Research on Cancer (IARC) classifications. Developmental toxicants were weighted based on hazard scores (outlined in EPA’s TSCA Work Plan Chemicals: Methods Document) that were assigned based their acute no observed adverse effect levels (NOAELs) derived from the Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles and EPA’s Integrated Risk Information System (IRIS) assessments. Asthma chemicals were weighted based on log-transformed Reference Concentrations (RfCs) from EPA’s IRIS assessments. These weighted factors roughly capture the relative degree to which the toxicity of a chemical contributes to a certain health outcome, amongst the full set of chemicals that are linked to it. These weightings are used

in assigning ‘Weighted Risk Levels’ of Lower, Higher or Highest on the map pages relating to Cancer, Developmental Harm, and Asthma.

Chemical releases by facility were reported in unique chemical release per facility entries. Summarized data such as health harm-specific summary info and demographic buffer data was reported at the facility level. Each entry delineates the following:

- The raw pounds released of each chemical (2018-2022). For PFAS chemicals, releases for each year for each of 180 PFAS were summed to represent PFAS as one chemical group.
- The annual minima, maxima, and the sum of pounds of chemicals released during 2018-2022 for each facility (calculated as the sum of each chemical’s highest one year release between 2018 and 2022);
- A facility percentile rank and weighted risk level estimated according to the sum of pounds released;
- The same metrics as above but subsetted only to chemicals linked to each health harm of interest (cancer, developmental harm, and asthma);
- A variable denoting a count of how many of the three health harms of interest the chemicals released at a facility are linked to;
- Demographic metrics obtained from the prior steps, for the 10 km buffer around facilities and county-wide values to provide a representative control group for comparison, in terms of roughly gauging over/under-representation of vulnerable populations in proximity of facilities;
- Summary of number of facilities, chemicals, and total sum of releases (calculated as the sum of each facility’s sum of each chemical’s highest one year release between 2018 and 2022) for each state and district;
- Estimates of the number of other facilities within a 10 km radius, and sums of each facility-specific sum of (annual maxima) pounds of chemicals released during 2018-2022. These are calculated as metrics provided in the action alerts to policymakers, and aim to characterize the cumulative nature of chemical releases.

Data were then exported and uploaded for online for display.

For more information, please contact Paige Varner, Healthy Communities Scientist, pvarner@edf.org

Environmental Defense Fund
257 Park Avenue South
New York, NY 10010

A vital Earth. For everyone.
EDF.org
T 212 505 2100