**Code Review/Process Notes for Annual Air Toxics Summary Report**

Step 0 Toxics Tools

Before running this script, we need to have the proper input files in the directory. The 1st step is to pull an AMP 501 report from AQS. To do this, you will need an AQS login. Once you have the correct credentials to access AQS, you will need to set up a few parameters to run the correct report.

From the AQS main page, go to the Retrieval -> Standard Report Selection tab.

From the Criteria Set tab, the report code should be AMP501. Filename field should be filled out with report name. Report output should be WORKFILE.

Graphical user interface, text, application

Description automatically generated

From the Data Selection Tab, complete fields for state code, county code, and site ID. That way the report will only pull from whatever sites we are interested in.

Graphical user interface, table

Description automatically generated

The AMP501 report will be a flat file (.txt). The script will “tidy” up this flat file and get it ready for analyses.

This script will also load the cross tables that have been manually built. The cross tables are used as references for the dataset and will help the analyses in future scripts. A key thing about the cross tables is to make sure they are maintained and updated when necessary. A good example is the Analyte\_Master\_LookupTable. In this table there is a column named oregon\_abc\_ug\_m3. This column needed to be updated with the current list of ABCs as they’ve changed since the table was first created.

To summarize, when running step 0, first pull the data from AQS and place the file in the correct place, in this case the input folder. Build the cross tables and place those in the correct folder. Also, in the input folder. Then run the script.

**Input**

Cross Tables

AMP 501 report from AQS

**Output**

N/A

Step 1 Air Toxics Load and Merge

Currently there are 2 functions of this script. One is to pull and merge data from AQS and the other is to pull and merge data from the Element repository. As a matter of process, we have decided to pull data from AQS and merge that with the cross tables that we have created. So, we will focus on that for the function of this script.

This is not the excel file that we will use to do our analyses. This script will create an input file for Step 3.

The purpose of this script is to load and merge the data we’ve pulled from AQS with the cross tables we built and placed in the input folder. When we do this and run the script, a new output table will be created named, “air\_toxics\_compiled\_”. A thing to note about the output file is that the date it was created will get added to the name. So, if we created the table on January 1st 2023, the file would be named “air\_toxics\_compiled\_01012023”

Step 2 Particulate Load and Merge

The purpose of this script is to pull in daily PM 2.5 data and compare the air toxics sample dates to see if there was wildfire smoke that were affecting samples. The script will produce an output table, /output/tables/particulates\_daily\_ (Excel) that has a column for wildfire\_impact. Filtering for “wildfire” will show the days that samples were impacted by wildfire smoke. Then match the sample dates with the wildfire dates to see what air toxics samples were affected by wildfire smoke. Calculate annual averages for all air toxics with wildfire days included and excluded from the data set. Then calculate the % change difference.

**Inputs**

Access AQS to generate AMP 501 report for the monitors and parameter number 88502 (PM 2.5)

Cross Table (previously built)

**Output**

/output/tables/particulates\_daily\_

Step 3 Air Toxics Analyses

**Inputs**

/output/tables/particulates\_daily\_

/output/tables/air\_toxics\_compiled\_

**Analyses in the Report**

1. Data completeness. There is a requirement to meet 85% of expected samples for a given quarter, based on the EPA sample schedule. Based on the start date of the site, there should be an expected # of samples. Due to voids and other reasons, we may not reach this 85% of expected samples. Analysis should tell us the number of expected samples and how many valid samples we have throughout the year. For the 2022 report, we need to look at the start dates of each site, the end date if applicable, and how many valid samples we have over the year.

Use air\_toxcis\_pivot\_ and find n\_obs for what year you want. Divide this number by the number of expected samples based on EPA sample schedule.

1. Annual averages compared to Chronic Cancer ABCs. Ambient Benchmark Concentrations have been developed and updated for several air toxics. We must compare each analytes annual average against its Chronic Cancer ABC. If the annual concentration is over 1, it represents a small increase in cancer risk to a population. List all averages that are over 1 for that analyte, at that site. Be careful to remove datasets with a high number of ND results.

Use air\_toxics\_pivot\_ and filter for site and year. Use column conc\_benchmark\_qmean\_ and filter for values that are above 1. List these values for the site in the report.

1. Comparison of annual averages that are over 1 to national average for those analytes. For this analysis, you will need to contact EPA to get their updated dataset. Contact at EPA for this would be Regi Oommen, [regi.oommen@erg.com](mailto:regi.oommen@erg.com) or Jeanette Reyes, [reyes.jeanette@epa.gov](mailto:reyes.jeanette@epa.gov). Create a table showing the national averages and how many times they are over the ABC.

1. Cumulative cancer risk and hazard index. All of the analytes monitored would contribute to excess cancer risk and/or chronic health risk. Even those that are below their ABC. For this analysis, we want to aggregate all the annual concentrations against their ABC for cancer risk. This will give us a total excess cancer risk for that site, for that year. In a similar way, we want to aggregate all concentrations against their chronic non cancer risk ABCs. For the non cancer risk, this will give us a hazard index. More information about hazard index can be found here [caononcancerfs.pdf (oregon.gov)](https://www.oregon.gov/deq/FilterDocs/caononcancerfs.pdf).

Use air\_toxcis\_pivot\_. Filter for site and year and use column conc\_benchmark\_qmean\_. Add all the numbers in this column to get total cancer risk for that site, for that year.

1. KPM trends. DEQ keeps track of 5 air toxics at the 2 NATTS sites to represent trends over the years. The 5 KPMs tracked are formaldehyde, benzene, cadmium, arsenic, and acetaldehyde. For this analysis, gather the last 5 years of annual averages for these 5 KPMs at the NATTS sites and plot a timeseries showing the trends over that 5 year period.

The easiest way to get this data is to download it from EPA. <https://www.epa.gov/outdoor-air-quality-data/monitor-values-report-hazardous-air-pollutants>. Download the year you want for the state of Oregon and filter for the KPM toxics to get the annual average for that year.

1. Comparison of individual samples to short term cancer risk. There is an ABC table built that lists the short term cancer risk ABC. We will use this to compare individual samples, to see if any are above this short term risk level. Individual samples above acute ABC are likely due to wildfire smoke but could be due to other reasons. If there are individual samples above short term risk levels, it’s a good idea to investigate further to provide details as to why.

To get results, use air\_toxics\_non\_cancer\_cao24hr\_exceed\_aqs. Filter action code column for |. Results will be in column best\_val\_ug\_m3.

1. Wildfire smoke impacts. If there is wildfire smoke in the area during a sample event, it is likely to raise levels of air toxics. This analysis would like to show what the annual average for air toxics would be with and without impacts from wildfire smoke. To identify wildfire smoke impacted days, we pull an AMP 501 report from AQS for 88502 (PM 2.5) to show daily levels of PM 2.5. We then match up the days of high levels of PM 2.5 (currently 25 μg/m3 or higher) with sample days. Calculate annual averages for all air toxics with wildfire smoke impacted days and without to see the % change difference. This should show us how much wildfire smoke is impacting air toxics annual averages.

Also, we can calculate excess cancer risk/health risks with wildfire smoke days included or not included in the annual average.

To find annual average and impacts to cancer risk with wildfire days included: air\_toxics\_summary\_aqs\_ column for conc\_benchmark\_mean

To find annual average and impacts for cancer risk with wildfire days removed: air\_toxics\_wildfire\_summary\_ column for conc\_benchmark\_mean and filter wildfire\_impact for none.

\*would like to add a column for % change between the two.

Questions I have about the table, air\_toxics\_compiled\_

1. In the column M, sample value, there are 0 values. What do those correspond to?
2. In the column M, sample value, there are NA values. What do those correspond to?
3. How is this table handling NDs?