ATSDR Substance Priority List (SPL) and Completed Exposure Pathway (CEP) Report Data through 2019

Cover Notes

www.atsdr.cdc.gov/SPL

Last updated June 12, 2020

Contents

Introduction 1
Substance Priority List Data 2
Metadata, Frequency, and Toxicity 2
Concentration Data for Water, Soil, and Air 4
Theoretical Daily Dose and Source Contribution 5
Exposure Columns 7
Additional Columns 8
SPL Version 2 (V2), as compared to Version 1 9
Initial Priority Lists 10
CEP Site Count Report (a.k.a. CEP Report) 10
Summary Statistics 11
Full SPL Spreadsheet Version History 13

Introduction

The Agency for Toxic Substances and Disease Registry (ATSDR) is part of the U.S. Department of Health and Human Services.

These Cover Notes explain the columns and data of the SPL and CEP Data worksheets. It is intended to be used with the "Full SPL Spreadsheet" Excel file (.XLSX) found at www.atsdr.cdc.gov/SPL

In the **Full SPL Spreadsheet**:

To see SPL data, use the **SPL Data** worksheet tab at the bottom of the spreadsheet window. For further information, see below or www.atsdr.cdc.gov/SPL

To see CEP Report data, use the **CEP Data** tab. Its columns are explained past the end of the SPL information (below). For further information, see below or www.atsdr.cdc.gov/CEP

Use Excel filters and sorting to see only, e.g., the most recent year (/Data/Filter on the Year column). For example, if you were to search the Name column for "Arsenic," you would find 20 rows in the SPL Data for all past years and versions. Filter the Year by "2019" and the Version by "1" to see only the most-recent official (Version 1) SPL.

This spreadsheet has many columns and is not ideal for printing; for more print-friendly versions, see listings on the web pages above.

There is also a **Summary Statistics** tab which includes supporting information for these lists; see below.

For more information on this spreadsheet, contact Mike Fay, Ph.D. (MFay@cdc.gov), ATSDR/DTHHS/ETB. Agency for Toxic Substances and Disease Registry, Division of Toxicology and Human Health Sciences, Environmental Toxicology Branch. 1600 Clifton Road Mailstop S102-1, Atlanta GA 30329 U.S.A.

Substance Priority List Data

Metadata, Frequency, and Toxicity

Version: There are two versions of the SPL: Version 1 (V1) is our traditional SPL focused on substances of greatest public health concern (to prioritize Toxicological Profiles), begun in 1991. Version 2 (V2) began in 2013 and focuses more on public health impact. The notes on this cover page discuss V1 of the SPL, and often apply to V2 data with only slight differences. Then V2 itself is specifically described after the V1 notes (below).

There is also a version “0” in this column for our pre-1991 lists (before our computerized algorithm) to let these be filtered in or out easily. See “ATSDR SPL Dates and Notices” on the SPL Resources page for more on this old version, and the history of the SPL.

Year: The year of record for each list. This is not always the year the list was generated or published if it was delayed; see the Summary Statistics tab and/or the "ATSDR SPL Dates and Notices" PDF on the SPL Resources page for run dates and Notice dates.

Line: This number is similar to the Rank, except the Line number increments for every single row. (Unlike Rank, it doesn't repeat if the Total Points are equal for some rows.) The Line number is a unique number for reference purposes, or to re-sort to the exact order the substances were originally presented in. (Substances are sorted based on decreasing Total Points, then Sort Name.)

Rank: The rank of this chemical for the current SPL release (current as of the Year column). Substances were sorted on the Total Points column, and the substance with the highest Total Points received the lowest rank (1). Rank is repeated if Total Points are the same (those substances tied for that rank). Chemicals often have tied ranks farther down the list, where substances are data poor and have little information to distinguish them. Also, two large groups (negligible toxicity and Petroleum Exclusion substances) have been forced to the bottom of the list with very low Total Points.

CAS RN: The Chemical Abstracts Service (CAS) Registry Number for this substance.

CAS RNs are a popular way to identify substances, but they have some ambiguities. For example, there is an overall CAS for PAHs, but also numerous individual PAH molecules and CASs. Likewise, there are individual PCB molecules with CASs, but also CASs for Aroclor mixtures of those same molecules. Fuels and other complex mixtures have their own CASs, as well.

We use CAS RNs instead of groups like “mercury compounds” because analytical results and site documents typically identify specific compounds. Decisions to group substances together and lose this specificity should be done after individualized data has been summarized (herein). Also, if we had pre-packaged our substances into groups, you would not be able to see specifics on, e.g., individual PCB congeners as they appear in documents; there would only be one condensed PCBs summary record for them all.

If you want to group similar substances together, use the Tox Profile Number (see below).

Total Points: The total number of points for this chemical (adding Frequency, Toxicity, Source Contribution, and Exposure points). The highest theoretical Total Points is 1800 if one substance was highest on all four point columns (600+600+300+300). In two instances, Total Points is forced to a specific value: If it has negligible toxicity (Tox Points = 0), Total Points is forced to zero. If it is a Petroleum Exclusion substance, Total Points is forced to -1. Both of these effectively remove a candidate from consideration. For more information on the SPL algorithm, see the Support Document on the SPL Resources page.

Site Frequency: For V1, this is the number of EPA National Priorities List (NPL) sites for which ATSDR has finalized a document that discussed this substance, plus additional counts from NPL Site File records (see below). For V2, this is the number of sites and events at which it has appeared in finalized ATSDR documents (NPL or not). V2 is not simply “V1 plus more sites” because V2 drops some V1 records of lesser relevance (Site File records). See the V2 explanation at end of these V1 notes for more information.

In both versions, a substance must appear at 3 or more sites to be an SPL candidate. In practical terms, it must appear in an ATSDR site document (and/or Site File record for V1) and be abstracted into our science database for at least three sites. This should include all instances of contamination of concern, but the extent to which additional instances (below levels of concern) have been captured has changed over the years.

Initially (in the 1990s), most every mention of a substance in an ATSDR site document was abstracted, even low levels with no exposure. Starting in 2004 and certainly by 2008, the agency moved to abstract only substances in Completed Exposure Pathways at concentrations above a health guidance level. So, site frequency should not be considered an exhaustive count of every appearance of a substance at sites or even in ATSDR site documents. That said, there is a finalized, publicly-available ATSDR site document that supports all counts here (except V1 Site File counts).

Frequency Points: Number of points, based on its site frequency. This is a simple linear scale where the most-frequent substance is assigned 600 points (600 x Site_Frequency / Highest_Site_Frequency).

Toxicity: This is the value of the EPA Reportable Quantity (RQ) for this substance, or, if none exists, a "Toxicity/Environmental Score" (TES) developed by ATSDR using the same methodology. RQs are the number of pounds which must be reported if released; a lower RQ means that the substance is more toxic. Values range from 1 pound to 5,000 pounds. A value of 50,000 in this column means the substance is of negligible toxicity (>5,000). Also, Petroleum Exclusion substances are not assigned RQs or TESs. See the Methodology for Toxicity/Environmental Scores PDF and the SPL Toxicity Values XLS on the SPL Resources page for more information.

Toxicity Points: These are assigned based on Toxicity as follows:

RQ 1 pound = 600 points
RQ 10 pounds = 400 points
RQ 100 pounds = 178 points
RQ 1,000 pounds = 53 points
RQ 5,000 pounds = 10 points
(RQ 50,000 pounds = 0 points) for substances of negligible toxicity

See the Support Document on the SPL Resource web page for more information.

Concentration Data for Water, Soil, and Air

Concentration data comes directly from ATSDR site documents; each datapoint here can be found in one of them. Data has been summarized into concentrations for the three major media types (water, soil, and air).

The following seven fields appear for each of the three major media. Two or more concentration datapoints were required before statistics were generated for a given major medium. Note that typically only a fraction (very roughly 10%) of concentrations are from Completed Exposure Pathways (CEPs). Further, concentrations are not filtered relative to health guidance levels. As a general rule across this large historical dataset, the majority of concentrations are below levels of concern. Also, sometimes multiple documents for a given site repeat earlier analytical findings, resulting in duplicate data. It would have been difficult to control for this compared to simply abstracting what each document says. Thus, concentration statistics should be used with caution because some duplication and some summarization has occurred. However, it is still perhaps the best publicly-available aggregate concentration information for substances at major U.S. hazardous waste sites.

GMMC: Geometric mean of maximum concentrations for the three major media. ATSDR's science dataset summarizes its site documents and most of the concentration data was collected as the maximum concentration found for approximately 30 specific categories of environmental-medium pathways (types of air, water, soil, and other media; see Data Count, below). Only the maxima for these specific environmental media were collected for each ATSDR site document. Also, up to the year 2004 ATSDR generally abstracted every maximum concentration for each type of medium appearing in an ATSDR document, regardless of exposure status or level of concern. Due to decreased funding, starting in 2004 and fully by 2008, data were usually only collected for Completed Exposure Pathway contaminants that were above health guidance values (MRLs and RfD/RfCs). The net result is that only a small percent of the concentration data points are now collected each year as opposed to what was collected up to 2004, but they are the most important contaminant data points.

GSD of MCs (2011 and later): Geometric standard deviation (of the sample) for the preceding geometric mean of maximum concentrations. A GSD of 1 (the antilog of 0) means that all concentrations had the same value.

A word of caution if you are converting the units of a GM and GSD: If you wanted to convert, say, water concentrations from mg/L to ug/L, you would multiply the GM by 1,000, but *not* the GSD. The geometric standard deviation is a unit-less factor perhaps better called the **GSD factor**. For more information, see, e.g., https://en.wikipedia.org/wiki/Geometric_standard_deviation. As per Wikipedia, in contrast to the arithmetic mean and SD,

When using the geometric SD factor in conjunction with the geometric mean, it should be described as "the range from (the geometric mean divided by the geometric SD factor) to (the geometric mean multiplied by the geometric SD factor), and one cannot add/subtract the "geometric SD factor" to/from the geometric mean.

Median of MCs (2011 and later): Median concentration of the maximum concentrations. This is a standard arithmetic median; if there were an even number of datapoints, the two central data were averaged with a simple average (not a geometric average).

Min and Max of MCs (2011 and later): These data are provided on an "as is" basis and should be used with caution. The minimum values mostly reflect the limit of detection. Consider that there will always be plenty of instances of no concentration (non detect) as well, but our dataset does not capture such events. (We are also not able to say how many times a substance was tested for but not found.)

Likewise, maximum values can be due to exceptional circumstances, such as drilling into a clump of near-pure substance found only at one place on one site.

Finally, ATSDR's science database was designed to give an overview of site contamination, not detail all site data comprehensively (an expensive proposition). Acceptable data accuracy was set at 0.5%. Among the hundreds of thousands of contaminant records, there have been a few concentrations of "greater than pure" and other outliers. Given the history of the data and current level of funding, it is not feasible to fix these problems.

In summary, these two fields are especially sensitive to problematic outliers and should only be used for informal observations.

Data Count: This is the "N" for the number of maximum concentrations used for the statistics. For practical purposes, it is also the number of pathways in which the substance was found in ATSDR documents, and subsequently abstracted into our science dataset. (As explained above, the amount of data abstracted per document decreased significantly after 2004.) When data is collected, only one maximum is allowed for each of approximately 30 types of submedia (surface soil, subsurface soil, drinking water wells, surface water, monitoring wells, concrete, etc.). Therefore, while this number approximates a pathway count, it is more accurate to say it is the number of submedia types that were contaminated. There could have been a number of measurements of surface soil, but only the maximum concentration will be entered for that submedium (surface soil) from that site document.

Because it is rare for there to have been multiple, distinct, identical submedia pathways of concern (e.g., surface soil with concentrations of concern both on the south side of a hazardous waste site and its northeast corner), this count is usually, but not always, equal to the number of pathways of concern. Historically, half of NPL sites with public health assessments (larger sites) in our contaminant dataset did not have a single completed exposure pathway. And of those that had at least one CEP, the bulk only had one.

Site Count: The number of NPL sites at which the substance was found, for the previous concentration statistics. Since a substance can appear in more than one pathway at a site, the Site count is equal to or less than the data (a.k.a. pathway) count.

Theoretical Daily Dose and Source Contribution

TDD: This is the theoretical daily dose for the compound, composed by applying exposure assumptions to the GMMCs. See the Support Document for more information. It is important to note that ATSDR neither states nor believes the public was exposed to these levels, especially given that we abstracted the maximum concentration found at the entire site for each submedium. The SPL, including the TDD, is only designed to prioritize which substances found at sites shall be candidates for Toxicological Profiles. We are using a simple, straightforward algorithm atop summary data, to assist in discriminating substances.

TDD Site Count: The number of sites at which concentration data for the TDD (i.e., for all three media) were found. Since a substance can be found in more than one type of medium at a site, the TDD Site Count is equal to or less than the number derived by adding together site counts for the three media, and it is greater than or equal to the largest of the site counts for air, water, and soil. For research purposes, this number shows how many sites had two or more usable concentrations for any of the three major media.

The difference between this value and the Site Frequency count is that our dataset draws on several sources, not all of which have concentration information, nor is a stated concentration always usable for SPL purposes. Overall, only a third of sites' records have suitable concentration data for V1. (It includes the Site File records, none of which have concentration data. See below.) Except for Site File records, most of the difference is due to instances of contamination that were not of concern (in eliminated pathways or even just trace contaminants) or poorly quantified (verbal anecdotes or out of range of the analytical method). The SPL is more focused on prioritizing chemicals in general (especially data-poor ones) than focusing on contamination of concern. The V2 results fare better relative to substances of concern, largely because V2 excludes the Site File data (see below).

SC: Source Contribution. This approximation is the TDD divided by the Toxicity value. In other words, substances with lower toxicity (and a higher Toxicity RQ-pounds value) will have their TDD reduced more than substances with higher toxicity (and lower Toxicity value). Thus, for purposes of prioritization, highly-toxic substances (Toxicity=1) do not have their TDD reduced. As with other SPL components, this is for scaling purposes only, and should not be taken as a statement of any actual exposure.

SC Points: Source Contribution Points. This is the first half of the exposure component of the algorithm. There is a minimum of 0 and a maximum of 300 points available. Scoring is based on logarithmic scaling of the SCs and, since 1995, is based on the geometric mean (GM) and standard deviation (GSD) of all available SCs. This includes high and low cut-offs at two GSDs from the GM of the SCs. Thus, an SC at the GM of all SCs receives 150 points. SCs that are at or below -2 GSDs from the GM receive 0 points, and at or above +2 GSDs receive 300 points. This keeps most (95%) of the SCs within the scoring range (0 to 300 SC Points) and the GM in the middle of it, and prevents the most extreme value from moving the center of the point-score distribution away from 150. Petroleum Exclusion and negligible toxicity substances do not receive an SC nor any SC points. For more information, see the Support Document.

GM, GSD, and other statistics for the SC computation are available on request, but they can also be directly computed (and point scoring confirmed) from the data presented herein. Simply make a column with log values of the SC, then derive the average and standard deviation of the sample from these values, for a given year and SPL version.

Exposure Columns

These columns are, in effect, the number of different exposure pathways (out of the approximately 30 media categories) where the substance was mentioned in ATSDR site documents. Importantly, this is not the number of people exposed. Nor is it for pathways with the substance above a particular concentration. It is simply a count of pathways the substance was found in, at any concentration. For additional information on pathway counts, see the notes for Data Count (above).

The "**Exposure to Contaminant**" column (Category 1) counts the number of times that a substance was explicitly mentioned as being in a Completed Exposure Pathway (CEP) in an ATSDR site document. If a substance has counts here, it is given from 200 (count of 1) to 300 (highest count in this category) exposure points. This column is similar to the CEP Site Count Report, except that this is the number of pathways in which the substance appeared, not the number of sites. (A substance can appear in more than one pathway at a site.) Also see the Site column description, below.

The two "**Exposure to Media**" columns (Category 2 and 3) count the number of times that a substance might implicitly have led to exposure at a site. "Implicit" means that an ATSDR site document might have stated that people drank groundwater via wells, and it also stated that the substance had been found in groundwater - but the document did not explicitly state that the substance was found in a groundwater completed exposure pathway, per se. (For example, it was in groundwater far from a well that was used.) The Implicit Exposure columns are essentially artificial constructions that lend more discriminatory data points particularly to data-poor substances with little or no exposure data. "**Potential Exposure to Media**" (Category 3) means that the document mentioned potential exposure to the medium pathway, whereas there was clear exposure for Category 2. As with Category 1, Category 2 can receive from 100 to 200 points, and Category 3 can receive 1 to 100 points.

The design of the Exposure to Media columns has occasionally been adjusted to increase or decrease the number of counts derived, and therefore the amount of discrimination between chemicals from these data. This has occasionally led to some very high counts, especially in the early 1990s. However, in any given SPL year, the same algorithm was applied to all chemicals in that year, so that the desired relative rankings were still achieved. In 2011, a more robust scheme was implemented.

Points assigned based on counts in the exposure categories are not cumulative. If a substance has a count in Category 1, it receives 200-300 points (per the above), and does not receive points for any Category 2 or 3 counts. Likewise, if there is no Category 1 count but there is a Category 2 count, it receives 100-200 points, but nothing for Category 3 counts.

The design of the algorithm effectively makes high scores in Category 2 or 3 very improbable. This is because a substance that is found in numerous media pathways at numerous sites is also likely to have occurrences of exposure to the substance (Category 1). Thus, its Category 1 score prevails over its Category 2 or Category 3 score, as discussed. Due to this "masking" effect, only substances with exposure via media at a few sites have Category 2 or 3 scores that are not masked by Category 1 occurrences. Therefore, exposure point scores based on Category 2 or 3 data alone are on the low end of the range of points available for those two categories. This effect on the point score is appropriate, because actual documented existence of exposure to a substance (Category 1) is, of course, a considerably more reliable measure of exposure than indicators based solely on the inferred possibility of exposure via media (Categories 2 and 3).

The "**Potential Exposure to Contaminant**" category (Category 4) is no longer used. It counts the number of times that a substance was explicitly mentioned as potentially being in an exposure pathway, but there was no known actual exposure. This Category was part of the initial SPL design (1991 to 1993) but, as can be seen in the data for those years, it was almost never actually used for exposure points. This is because it was rarely used in the source documents and, if it was, there was almost always also an actual exposure (Category 1), which then masked this category. Thus, it was removed to lend more discrimination to the exposure columns that did contribute useful information. The exposure-point scoring scheme was also slightly different when this category was used. It was originally Category 2, but has since been dropped to 4 to move obsolete information out of the way in this current spreadsheet. In the ***old*** scoring scheme (1991 to 1993), Category 4 (then Pot. Exp. to Media) received 1-75 points, Category 3 (then Exp. to Media) received 75-150 points, Category 2 (then Pot. Exp. to Contam.) received 150-225 points, and Category 1 (Exp. to Contam.) received 225-300 points.

For all four of the Exposure columns, there is an accompanying **Site count** column. A substance may have appeared in more than one pathway at a site, so the site count is always less than or equal to the exposure (pathway) count. Note that the V1 Exposure To Contaminant (Category 1) Sites column is equal to ATSDR's CEP Site Count Report for that substance at NPL sites, for a given SPL Year. Occasionally there are extremely minor differences due to data entry between the time the SPL was run and the CEP Report was run; they are run in quick succession, but it is still conceivable a little data was entered in the meantime. Interested users are encouraged to compare the CEP Report because it includes additional information not found in the SPL concerning both non-NPL sites and substances without specific CAS RNs that are difficult to identify (process wastes, emergency-event "acids not otherwise specified," etc.). See the CEP Site Count Report section (below) for more information.

Exposure Category Used indicates which category (1-4) was used for the Exposure Points for that substance.

Exposure Points: The number of points assigned to this substance, based on its exposure-count columns. Values can vary from 0 (no exposure) to 300 (for the substance with the highest Category 1 count). See above and the Support Document for more details.

Additional Columns

Sort Name: This column can be used to sort substances into a more user-friendly order, regardless of the particular form of the name found in the Substance Name column. For example, the Substance Name might be 1,3,5-Trichlorobenze, but its Sort Name starts with Trichlorobenzene, so that other trichlorobenzenes sort together with it. Sort Names are only provided for sorting purposes and can be incorrect as substance names, per se.

CAS RN with leading zeroes: This is the same as the CAS RN column, but padded with leading zeroes to be 11 characters in length. Useful for sorting on the RN as a number, having right justification by default or when not otherwise available (such as a TXT or CSV), or to have column-aligned RN digits.

Tox Profile Cover Date: Starting in 2017, we list the cover date of the Toxicological Profile in which this substance appears (if any). These dates are only good for the year and month (Excel dates must have a day of month). This information is current as of when that year's SPL or CEP Report was released. It will be out of date for older lists. (For example, some Profiles were updated after 2017, but the 2017 SPL rows will only show cover dates through 2017.)

This field may be blank for infrequent variations (salts, etc.), certain related substances, or poorly-defined forms of a chemical (NOSs and HZs) that were not specifically mentioned in a Profile. Stated another way, we do not expect false positives here (if we say it's in a Profile, it will be), but there may be false negatives (we may not have cited every form of a substance). We welcome suggestions for improvement here.

Tox Profile Number: Starting in 2017 (2013 for CEP), we list the Tox Profile Number (or TP Number) this substance appeared in, if any. (Each TP Number should have a Cover Date, and the same caveats apply.) This is a unique number which can be used to group substances from the same Profile together. The values are composed as follows:

The two-letter prefix in the data means:

TP: a normal CERCLA Tox Profile
DD: a set funded by the Department of Defense (DoD)
DE: a set funded by the Department of Energy (DoE)

The year within the TP Number is not the year it was published; it is the year this Profile was initially designated for development (when that set was created). Typically, Draft Tox Profiles are released a year later, and Final Tox Profiles two years later. Some Profiles take longer for various reasons. See the Cover Date for the actual date of the document.

The number after the slash is a sequential number for the Tox Profiles in that set. Example: if there were 10 Tox Profiles in a set, this number would go from 01 to 10.

If this data is sorted on Tox Profile set numbering, many gaps will be seen. Often, DoD and DoE Tox Profiles, and certain Profiles developed for, e.g., emergency responders, were not driven by the SPL and therefore are not represented in the SPL data. Furthermore, a very large number of Tox Profiles have been updated and re-released in a subsequent set. Only the latest TP Number for a substance (as of each List) is shown.

SPL Version 2 (V2), as compared to Version 1

Starting in 2013, ATSDR began producing an alternate version of the SPL. The algorithm is exactly the same, but some of the data going into the algorithm is different. Here is how the data differs from the original, traditional List (version 1 or V1):

1) V1 includes an archival record set known internally as the "Site File" data. This is a fixed set of 88,582 instances of contaminants found in an environmental medium at NPL sites (see below for more dataset statistics). It does not have concentration or exposure information, nor is it necessarily of concern. In the early 1990s, it was gathered directly from our EPA source "site file" documents as a test that ATSDR documents and data abstraction were robust. Later, it was used simply for additional discrimination (prioritization) in site counts, so it might help data-poor SPL substances. But we exclude it from V2 in order to focus on more relevant data. Stated another way: Because ATSDR staff extract contaminants of concern from source EPA data into the documents we write, if something appeared in the Site File but did not appear in ATSDR document on that site, it is because it was not of concern. EPA technical site documents sometimes have long lists of analytical results, many of which might not be of public health concern.

2) V2 includes data from non-NPL sites, as well as the same NPL sites in V1 (but not Site File data). However, because NPL sites tend to be larger, data-rich sites, the non-NPL sites and events added for V2 are generally for relatively small or very small sites, events, and emergencies (sometimes even phone calls). The additional ATSDR-specific sites and events typically have far fewer substances per site/event than NPL sites. There are a few notable exceptions (very large sites that are not on the NPL).

Because of these two changes, V2 results are based on fewer contaminant records but ones of higher public health relevance to ATSDR: in 2013, there were 181,682 V2 contaminant records versus 200,365 in V1 (including its 88,582 Site File records).

Yet we found that, overall, there was not much change in the top-most chemicals. This is because the substances that are highly frequent stay that way; they're also more liable to have Category 1 exposure counts, and, of course, substance toxicity stays the same. V2 still has (non-Site File) NPL records, and they still represent much of the data.

Because there were not large differences, we kept V1 as the official SPL after having experimentally programmed V2. We decided it was not constructive to explain that we had improved the underlying records used, but it didn't change the top-most substances. Still, we provide V2 in this download for users that may prefer data more focused on contamination of public health concern. You will find that site counts are generally lower, but there is more concentration and exposure data.

Note that V2 more closely approximates the Completed Exposure Pathway (CEP) Site Count Report, especially its Exposure Category 1 counts. However, the CEP Report also includes poorly-identified HZ substances (see below).

Initial Priority Lists

In ATSDR's first years (before its science database was established), Substance Priority Lists used an expert panel weight-of-evidence approach, as described in the Federal Register notices (FRNs) for the years 1987 to 1990 in the “SPL Dates And Notices” PDF on the SPL Resources web page. Instead of the point-scored algorithm used in subsequent years, they were published in groups each year, as shown by the following fields. The grouping scheme was discontinued, and these prior rankings replaced, with the advent of the 1991 SPL:

Priority Group: As stated in the FRNs, these are the prioritization groups for the pre-1991 listings.

Priority Group Order: The order of the substances within each group, as listed in the FRN. Sometimes more than one chemical would appear on a single line of the FRN, for closely-related substances such as DDT, DDE, and DDD. Use this value to have the same order and logical grouping found in the FRN.

CEP Site Count Report (a.k.a. CEP Report)

The Completed Exposure Pathway (CEP) Report is an estimate of the number of hazardous waste sites with exposure to a chemical, in the United States. Strictly speaking, it tallies how often a chemical was found in a CEP in ATSDR site documents. Counts include EPA National Priority List sites as well as other sites for which ATSDR has written particular documents. It is not an exhaustive list; states, counties, cities, and other entities track many other hazardous waste sites.

This is a tally of the number of times the substance appeared in a pathway to which human exposure was likely, according to ATSDR site documents. It does not imply toxicity. This list also does not count or estimate the number of people exposed.

Fields for this Report are usually identical to the descriptions shown above for SPL, taking into account any notes about the CEP Report therein. Here are fields that need additional explanation:

Rank: From 2011 onward, Rank in the CEP Report is based on 1) decreasing number of All Sites/Events, then 2) decreasing order of NPL events. This is because non-NPL sites often (but not always) have fewer instances of contamination than NPL sites. If both of these stay the same for successive records, these records have the same Rank.

Prior to 2011, it is only based on decreasing number of All Sites/Events.

"All Sites/Events" versus NPL Sites: EPA determines which sites are National Priorities List (NPL) sites. ATSDR performs public-health follow-up at these and other sites.

"All Sites/Events" includes NPL sites, as well as all other sites of concern to ATSDR. Strictly speaking, this is defined for practical purposes here as sites or events for which contaminant information has been abstracted into ATSDR's science database. Data comes from ATSDR documents such as public health assessments, health consultations, and Epidemiological Investigations, but does not include more limited agency activity records such as Technical Assists.

Event versus Site: Some documents have been written for minor events and are not tied to specific, defined sites (emergency events, spills, etc.). These events are included in the "All Sites/Events" count if a health consultation was written for the event. NPL sites are well defined and, therefore, are never events.

Pathway Counts (2015 and later): The number of environmental media pathways (CEPs) in which the chemical was found. ATSDR's science database summarizes by type of environmental media, which can underestimate the number of pathways at a site. Also, if there are multiple documents on a site that discuss the same pathway, there will be duplicate counts (an overestimate of pathway counts, but not site counts). The net result is that the pathway count is inexact, but higher aggregate values logically indicate more exposure pathways. Use the site count for an exact number. For more information, see the "Data Count" and "Site Count" explanations for the SPL concentration Data (above).

CAS RN: The Chemical Abstracts Service (CAS) Registry Number for this substance. "Pseudo-CAS RNs" starting with HZ are internal ATSDR numbers used to document process wastes and other non-specific instances of contamination that are of concern. While some of the HZ substances are quite imprecise, the substances cited were still of concern and mentioned in a health consultation for, e.g., emergency events, spills, and other situations. Therefore, they help inform users of this list as to which substances are cited as being of concern, even if they were poorly identified in the event.

Summary Statistics

These summary statistics started being collected in 1997, when the CEP Site Count Report was first produced. Not all datapoints are available for 1997. No list was produced In 2009, while the agency transitioned to a new science database.

This information can be used in two ways:

1) As a measure of the amount of information available to ATSDR in its science database for each List year.
2) As a denominator for data from the SPL and CEP Reports to make statements such as "x percent of NPL sites with ATSDR activity had arsenic present in 2005."

Both the SPL and CEP Site Count Report are run in succession on the same day; these statistics apply to them both.

The numbers are only produced for, and applicable to, V1 and the CEP report. They are only intended as general baseline contextualization values. V2 is produced at the same time from some of the same data as V1, and V2 values follow the same tendencies, but are not identical. (See discussion of Version 2, above.) Summary Statistics for V2 might be able to be reconstructed, if needed.

The **Run Date** is provided for anyone interested in comparing how much the SPL data changes over time. Here, “run” means when the List was output or generated from the underlying data.

For a variety of reasons, these lists were not run on the same calendar day of the year, for each SPL run. Thus, in order to properly compare changes over time, one must know the exact day each of the Lists was run. Data is dynamic (constantly entered) and will have been captured up to that date. For the record, site data entry often lags behind document cover dates, and can take from a few weeks to a few months to be entered. A higher percent of documents are released at the end of ATSDR’s fiscal year (Sept. 30). Information from this rush of documents may not be fully entered into the science database until the beginning of the next calendar year.

“Events” include ATSDR emergency events and other incidents involving hazardous substances, often very minor in nature, maybe even just a phone call. Thus, there are far more ATSDR sites and events than there are NPL sites. For a general measure of the ratio of larger ATSDR sites to events, compare the number of public health assessments (PHAs). Generally, larger sites tend to have at least one PHA, but minor sites may only have a health consult (HC), and events may have neither PHAs or HCs. NPL sites are defined by EPA and therefore will never be an ATSDR event.

The Summary Statistics rows (1 to 6) are not necessarily subsets of the row above them. Thus, row 3 (Site/Events with CEPs) appears to have a much greater increase in CEPs at All Sites/Events than seems possible given the increase for NPL sites - if half of NPL sites with PHAs have a CEP (e.g., 802/1621 for 2011), why do most of All Sites/Events seem to have one (1686/1831)? This would be comparing apples to oranges; row 3 is not versus row 2, it is versus row 1. While most NPL sites are large and have a PHA (or soon will), the universe of All Site/Events for row 3 is much larger and more approximates the 6096 Sites/Events on row 1. Many of the non-NPL sites and events don't and won't have a PHA, but instead come from numerous smaller documents. They also finalize more quickly, so there is less of a draft-to-final "gap."

There are some reductions in counts from 2007 to 2011 when ATSDR transitioned to a new science database because the previous algorithm included draft data, if the site document was not yet finalized. Such data is problematic because not every datum in a draft document makes it to the final document. We decided to use only data from finalized site documents in the algorithm when it was reprogrammed for our new science database, Sequoia. It would be too much effort to reconstruct how the data was at all points in the past in order to produce completely standardized Summary Statistics using only finalized

data, retroactively. Also there were perforce some minor changes in the algorithm because the new system does not, e.g., define environmental media exactly the same way that the old system did.

These Summary Statistics are intended to show general context.

Full SPL Spreadsheet Version History

The Full SPL Spreadsheet was first released in 2011. ATSDR's new science data system (Sequoia) provided more robust information, and we also decided to collate all past SPL data into this spreadsheet, for convenience. The original version of the Full SPL Spreadsheet had extensive text notes on its Cover Notes and Summary Statistics worksheets. The intention was to ensure that the notes stayed with the data.

In 2019, we put the text notes into this document, for increased legibility. We also made numerous small edits for clarity. The SPL is now being presented as the Excel spreadsheet itself and a separate PDF of text notes, combined into a compressed zip package.