How to Use *CDC Wonder*

CDC Wonder is a search engine to selected CDC datasets – [https://wonder.cdc.gov/](https://wonder.cdc.gov/)

These are “Data Summary Descriptions” of those datasets – [https://wonder.cdc.gov/DataSets.html](https://wonder.cdc.gov/DataSets.html)

This is the CDC Wonder homepage –

![CDC Wonder homepage](image)

You’ll see that there are three top tabs in it: the default of *Wonder* Systems but also the equally valid choices of entry into CDC Wonder data of *Topics* and *A-Z Index*. Using the tab of *Topics* is my preferred point of entry into CDC Wonder; the main use that I make of the *A – Z Index* is as a shortcut for going to a dataset whose name I know.
This is *Topics* – [https://wonder.cdc.gov/WelcomeT.html](https://wonder.cdc.gov/WelcomeT.html)

This is only the top of this screen; scroll down to see more –

---

This is the *A–Z Index* – [https://wonder.cdc.gov/WelcomeA.html](https://wonder.cdc.gov/WelcomeA.html)

This also is only the *top* of this screen; scroll down to see more –
Therefore CDC Wonder in effect has 3 very different entry points into it depending on what your search is and how you like to work with data.

The term WONDER here = “Wide-ranging Online Data for Epidemiologic Research”. Here is what CDC says about it –

CDC WONDER, developed by the Centers for Disease Control and Prevention (CDC), is an integrated information and communication system for public health. Its purposes are:

1. To promote information-driven decision making by placing timely, useful facts in the hands of public health practitioners and researchers, and
2. To provide the general public with access to specific and detailed information from CDC.

With CDC WONDER you can:

- Access statistical research data published by CDC, as well as reference materials, reports and guidelines on health-related topics;
- Query numeric data sets on CDC’s computers, via “fill-in-the blank” web pages. Public-use data sets about mortality (deaths), cancer incidence, HIV and AIDS, tuberculosis, vaccinations, natality (births), census data and many other topics are available for query, and the requested data are readily summarized and analyzed, with dynamically calculated statistics, charts and maps.
This is the CDC help guide to using CDC Wonder –
https://wonder.cdc.gov/wonder/help/main.html#What%20is%20WONDER

It’s a standard online guide, but is too short to be anything other than basic searching for users who don’t have a background in public health topics. It’s best to have a background or to be willing to research these health topics to get the most out of CDC Wonder.

There are several datasets in CDC Wonder that include this warning –

**Data Use Restrictions:**

The Public Health Service Act (42 U.S.C. 242m(d)) provides that the data collected by the National Center for Health Statistics (NCHS) may be used only for the purpose for which they were obtained; any effort to determine the identity of any reported cases, or to use the information for any purpose other than for health statistical reporting and analysis, is against the law. Therefore users will:

- Use these data for health statistical reporting and analysis only.
- For sub-national geography, do not present or publish birth or death counts of 9 or fewer or birth or death rates based on counts of nine or fewer (in figures, graphs, maps, tables, etc.).
- Make no attempt to learn the identity of any person or establishment included in these data.
- Make no disclosure or other use of the identity of any person or establishment discovered inadvertently and advise the NCHS Confidentiality Officer of any such discovery.

Before you enter the database you have to click this –

*By clicking the “I Agree” button I signify that I will abide by the terms of data use stated above and understand the sanctions and legal penalties for violation of these terms of use.*

I Agree

But in practice sensitive personal information in these datasets is suppressed anyway; example –

![Linked Birth / Infant Death Records, 2007-2014 Results](image-url)
Because many CDC Wonder choices link to databases on the main CDC web page I’m going to emphasize those that don’t, many of which are very involved.

Let’s do a **simple search** that doesn’t make too many data option choices: this is a search in “Cancer Statistics Data: Cancer Incidence” by these choices: Female, Black or African American and White, for Arkansas, Kansas, Missouri, and Oklahoma, 2010 through 2014, age 50-59, and cancer incidence in the small intestine.

Here are my deliberately simplified search choices –

1. **Organize table layout:**
   - Group Results By Ethnicity
   - And By States
   - And By Race
   - And By None

2. **Select location:**
   - Click a button to select locations by State, Region, or MSA.
   - States
   - Regions
   - MSA
   - States and Puerto Rico
   - Arkansas

3. **Select year and demographics:**
   - **Hint:** Use Ctrl + Click for multiple selections, or Shift + Click for a range.
   - **Sex:** All Sexes, Female, Male
   - **Age Groups:** 10-14 years, 15-19 years, 20-24 years, 25-29 years, 30-34 years, 35-39 years, 40-44 years, 45-49 years, 50-54 years, 55-59 years, 60-64 years
   - **Ethnicity:** All Ethnicities, Hispanic, Non-Hispanic, Unknown or Missing
   - **Race:** All Races, American Indian or Alaska Native, Asian or Pacific Islander, Black or African American, White, Other Races and Unknown combined
I’ve used my Control and Shift keys to make multiple choices in the above drilldowns.

Here’s the data –

You should read all “Messages”. I unclicked “Totals” but got the message about Totals anyway. Where data isn’t available we get zeros in our data table which by default aren’t shown, but should you want them to show you can choose them, along with additional display choices in “Quick Options”; the same for “suppressed” data.
If you want to add additional data and data display choices there are many of them. But you should consider every choice that you make carefully because every seemingly good choice could create unwanted data and/or data display results. **Spend the most time in “Request form” and think about the data you want and how you want them to be displayed; it’s easy to make mistakes here.** If the data and how they display isn’t what you want, keep returning to “Request form” until you get what you want or at least as much data as the rules of privacy don’t suppress. If you don’t make the right choices your data may not display at all because you haven’t created a field to display them in. *If you don’t get data and don’t see the red font “data suppressed” note in your returned data search you probably have a labeling problem somewhere in “Request Form”. Check everything twice, then again.*

The above example I call a simple search for CDC Wonder. But since some of its links go to simpler tools elsewhere on the CDC website, there are actually simpler types of searches. This is an example of a **quick search**. The following is another cancer data search within “Cancer Statistics Data” but is “SEER Cancer Query Systems: Cancer Prevalence Database: SEER 13” –

The following snips show how I got to this particular SEER 13 tool (there is more than one) –
1. I next click “SEER Cancer Query Systems” to get –

2. I next click “Cancer Prevalence Database” to get –
3. I next choose the “SEER 13” radial button and click to get it search engine; I’ve checked my search choices –

<table>
<thead>
<tr>
<th>Select Database</th>
<th>Step 1 of 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>○ SEER 18 Incidence - AA Rates for Additional Races, 2000-2015</td>
<td></td>
</tr>
<tr>
<td>○ SEER 9 Incidence - Crude Rates for White/Black/Other, 1973-2015</td>
<td></td>
</tr>
<tr>
<td>○ SEER 18 Incidence - Crude Rates for Additional Races, 2000-2015</td>
<td></td>
</tr>
<tr>
<td>○ SEER 18 Incidence - Trends of AA Rates for Additional Races, 2000-2015</td>
<td></td>
</tr>
</tbody>
</table>
### Selections

#### Statistic Type:
- [ ] Crude Prevalence Percent (SEER)
- [ ] Crude Percent Lower 95% CI
- [ ] Crude Percent Upper 95% CI
- [ ] Age-Adjusted Prevalence Percent (US 2000 Std (19 age groups - Census P25-1130))

#### Race:
- [ ] All Races
- [ ] White
- [ ] Black
- [ ] Asian/Pacific Islander
- [ ] Other Race

#### Sex:
- [ ] Male and Female
- [ ] Male
- [ ] Female

#### Tumor Inclusion Method:
- [ ] First Malignant Primary Only (Non-reported Assumed Malignant)
- [ ] First Malignant Tumor per Site and Years since Diagnosis
- [ ] First Malignant Tumor per Site in the Last 20 Years

#### Years since Diagnosis:
- [ ] 00 to ~05 yrs
- [ ] 05 to ~10 yrs
- [ ] 10 to ~15 yrs
- [ ] 15 to ~20 yrs

#### Site:
- [ ] Trachea, Mediastinum and Other Respiratory Organs
- [ ] Bones and Joints
- [ ] Soft Tissue Including Heart
- [ ] Skin excluding Basal and Squamous
- [ ] Melanoma of the Skin
- [ ] Other Non-Epithelial Skin
- [ ] Breast
- [ ] Female Genital System
- [ ] Cervix Uteri

#### Age at Prevalence:
- [ ] All Ages
- [ ] 00
- [ ] 01-04
- [ ] 05-09
- [ ] 10-14
- [ ] 15-19
- [ ] 20-24
- [ ] 25-29
- [ ] 30-34

---

**Go Back**  **Continue**
When I click “Continue” I get one more look at my search choices –

Cancer Query System: Cancer Prevalence Database

The Cancer Query System (CanQues) provides access to pre-calculated limited-duration prevalence estimates (percents and counts) and complete US prevalence estimated counts. Estimates are available by cancer site, race, sex, and age group.

Selections from Database: SEER 13 excluding Alaska, 23 Yr Limited Duration Prevalence on 1/1/2015 - Expanded Race/Ethnicity
Statistic Type: Crude Prevalence Percent (SEER)
Race: All Races
Sex: Male
Tumor Inclusion Method: First Malignant Tumor per Site and Years since Diagnosis
Years since Diagnosis: 10 to <15 yrs
Site: Breast
Age at Prevalence: All Ages

Define Results Table

No variables for stratification (all have single selections)

Number of Decimal Places for Statistics:
0.00005

Excludes whole numbers such as Counts

Go Back  View Statistics

And when I finally click “View Statistics” I get my data along with many notes about the data, all of which the user should read to either accept the search as done or to go back and reformulate their search strategy–
SEER 13 excluding Alaska, 23 Yr Limited Duration Prevalence on 1/1/2015 - Expanded Race/Ethnicity

Selections:
- Statistic Type = Crude Prevalence Percent (SEER);
- Race = All Races;
- Sex = Male;
- Tumor Inclusion Method = First Malignant Tumor per Site and Years since Diagnosis;
- Years since Diagnosis = 10 to <15 yrs;
- Site = Breast;
- Age at Prevalence = All Ages;

Results:

! 0.00164%

* Statistic not available by definition.
! Statistic derived from a limited age range.

Table Variable:

Notes:
- Statistics are provided by the Data Modeling Branch (DMB), NCI for research purposes only.
- Data Source: November 2017 Submission.
- Prevalence proportions are based on the standard SEER 13 registries, excluding Alaska.
- Populations from January 2015 were based on the average of the July 2014 and July 2015 population estimates from the US Bureau of Census.
- Survival Cohorts are based on:
  - Age at Diagnosis (00-03, 00-09, 70+ )
  - Years of Diagnosis (92-99, 00-04, 05-09, 10-14 )
  - Sex ( Male, Female )
  - Race/Ethnicity
    - For All Races, White, Black, and Asian/Pacific Islander: ( White, Black, American Indian/Alaskan Native, Asian or Pacific Islander, Other Unspec 1001 Unknown )
    - For Hispanic ( All races ) and Hispanic ( White ); ( White Hispanic, non-White Hispanic, Unknown Hispanic )
  - Sites ( Individual Site Recode with Mesothelioma and Kaposi Sarcoma )
- See Overview of Cancer Prevalence Statistics for more information about prevalence measures and the software used to generate the statistics.

Suggested Citation:

Alternative Formats:
- Delimited Data
We’ve done a **simple search** and a **quick search**. But we’re *not* going to do this **complicated search** that requires users to either already know much about the process and procedures of inoculation and/or for them to do research prior to formulating and executing their search because I don’t know enough about inoculation myself, have never had any customer interest here at ODL in it, and don’t have the time to research it now. This is **VAERS (Vaccine Adverse Effects Reporting System)** –

Look at this long, very long search engine screen and all the variables a user could choose for their search –
VAERS is complex enough that it even has the 13-minute Facebook video about it above. Among the variables users can choose are these selections—

* adverse effects after prior vaccinations
* medications at time of vaccination
* history/allergies
* symptoms
* vaccine characteristics: vaccine products, manufacturers, dose, lot #
* states and all territories
* event category, such as “Permanent disability”
* recovery status
* “serious” status
* vaccine administered by
* onset interval
* days in hospital
* vaccine purchased by (public, private, military...)

All users should note that each of the search terms is a link, so to get a definition of what “serious” status it above, click the link to get this –

Searchers can additionally search by terms using this tool –

Next there are 5 date-selection tools: Select report completed dates, Select report received dates, Select vaccination dates, Select adverse event onset dates, and Select death dates.
Finally there are these data display tools –

As complex as the search requirements are for VAERS, STD Cases (full name: Sexually Transmitted Disease Morbidity Data) is simple. These are my Request Forms choices –
* State
* United States, D.C. and outlying areas (Guam, Puerto Rico, U.S. Virgin Islands)
* Total Syphilis
* 2012, 2013, 2014
* All genders

After I got my data I sorted it in descending order: California and Texas were #1 and #2 in our nation for number of cases–
But then I sorted it by Rate per 100,000 –

<table>
<thead>
<tr>
<th>State</th>
<th>Count</th>
<th>Population</th>
<th>Rate Per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>District of Columbia</td>
<td>1,479</td>
<td>1,025,221</td>
<td>70.62</td>
</tr>
<tr>
<td>Louisiana</td>
<td>5,059</td>
<td>13,852,833</td>
<td>36.02</td>
</tr>
<tr>
<td>New York</td>
<td>18,614</td>
<td>58,872,515</td>
<td>31.02</td>
</tr>
<tr>
<td>Georgia</td>
<td>8,097</td>
<td>29,904,279</td>
<td>29.45</td>
</tr>
<tr>
<td>Texas</td>
<td>21,905</td>
<td>76,955,589</td>
<td>27.74</td>
</tr>
<tr>
<td>Florida</td>
<td>15,608</td>
<td>58,423,388</td>
<td>26.72</td>
</tr>
<tr>
<td>California</td>
<td>39,431</td>
<td>144,706,472</td>
<td>21.06</td>
</tr>
<tr>
<td>Maryland</td>
<td>4,079</td>
<td>17,742,191</td>
<td>22.99</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>2,475</td>
<td>10,897,266</td>
<td>22.71</td>
</tr>
<tr>
<td>Nevada</td>
<td>1,681</td>
<td>8,358,263</td>
<td>22.32</td>
</tr>
<tr>
<td>Illinois</td>
<td>7,881</td>
<td>38,039,525</td>
<td>20.49</td>
</tr>
</tbody>
</table>

Remember this about the listings of databases in any of the three entry-points into CDC Wonder: the links to go not only to main CDC.gov web pages, but also to the type of search engines that are unique to CDC Wonder –

**Births & Deaths**
- Births
- Infant Deaths
- Leading Causes of Death
- Mortality - Compressed
- Mortality - Occupational
- Mortality - Multiple Cause of Death
- Mortality - Quartely Provisional Estimates
- Mortality - Underlying Cause of Death

In the above list “Births,” “Infant deaths,” “Mortality-compressed,” “Mortality-Multiple causes of death” and “Mortality-Underlying cause of death” are all this type of CDC Wonder database –
But “Leading causes of Death” links out to WISQARS, “Mortality-Occupational” links out to NIOSH, and “Mortality –Quarterly Provision Estimates” links out the the National Vital Statistics System.

What are some of the more important databases in CDC Wonder? –

- **Cancer Statistics**
- **Disease Trends** – weekly reports
- **Mortality – Underlying Cause of Death**
- **National Notifiable Diseases Surveillance System (NNDSS)**
- **Online Tuberculosis Information System**
- **STD Cases**
- **Vaccine Adverse Event Reporting System**

But as I’ve said before in my other CDC webinars, if someone has a disease or adverse health condition—such as obesity—that’s *important* to them.
Let’s do a simple search –

*Mortality - Underlying Cause of Death*

This is what users have to read first and click “I agree” –

Only after a user accepts these conditions does the search engine load.

Notice these sentence from above: Deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal government violates 18 USC 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison, or both.

In the following example these are my search choices in Request Form; I accept all defaults other than what I change -

* State, Year
* All states
* 55 -64 years
* Male
* Black or African American
* 2014, 2015, 2016
* Sunday through Saturday
* Medical facility - inpatient, - outpatient or ER
* I21 – Acute myocardial infarction

Here’s the data –

![Underlying Cause of Death, 1999-2016 Results](image)

[web page continues through all states; links explain why data aren’t applicable for “Population” and “Crude Rate per 100,000”]

Let’s do another search, this time for two counties in Texas. Here are my search choices in Request Form –

* County
* Percent of total deaths
* Counties: Dallas and Denton
* All ages
* All genders
* Hispanic or Latino
* All races
* 2016
* Accept all defaults in Steps 5 and 6

Now how about choosing a geography other than “state”? We can choose “Census regions” or “HHS regions”. This is a more complicated search. For this example I’ve chosen all 4 Census regions and these search choices in *Request Form* –

* Census region
* Census regions 1, 2, 3, 4
* Micropolitan, non-core [rural geographies]
* 75-84 years of age
* Female
* Hispanic
* All races
* 2015, 2016
* Medical facility – inpatient
* J00-J98 All respiratory diseases
Now I’m going to change the “Select cause of death” to the more specific disease of “J09-J18” influenza-pneumonia –

Now I’m going to change the disease to be even more specific: can I get data for Streptococcus pneumonia? –

No! That’s too specific. So I’ll change my search to All categories (of geographies), all ages, all genders, all origins, all races –
Next I’ll change my Race to AIAN –

Next I’ll change genders to Hispanic or Latino Ethnicity, All Races –

This has been an illustration of how you can change your search strategy to get different data. The limitation that will frustrate you the most is the suppression of data to protect privacy. But remember this: CC Wonder is a primary source data search whose data in reported to CDC via state health agencies from sources within each state and territory. It’s data is as reliable as American FactFinder is for Census data.

One conclusion that all of you have now reached by now: there are many choices to make or not make in CDC Wonder, and unless you make good choices it won’t generate data. That’s why a CDC Wonder search isn’t a quick and easy search, but it is a primary source data search whose data comes from our nation’s state and territorial county health departments. This data is as reliable for health data as American FactFinder is for Census data.
Tools and Drilldown choices in CDC Wonder –
The easy way to review these is to simply work through the drilldown choices in “Request Form, but obvious tools are –
* the expand button in “Select year and month” and “ICD 10 Codes”
* the drilldown choices in “Group Results By” that includes injury and drug-alcohol cases
* the choice in “Select location” of States, Census Regions (4) and HHS Regions (10). Here are the maps of Census and HHS regions –

Census Regions –

![Census Regions and Divisions of the United States](image)
* There are differences in the 2006 and 2013 urbanization definitions as to whether which counties belong to which urban definitions. But if you’re searching older data, choose the older definition that better matches the data.

* In “Select Demographics” be careful of which age groups you choose because the age groups past the basic, default “Ten-Year Age Groups” have data suppression differences.

* In “Select Cause of Death” there are two expand tools: “Open” and “Open Fully.” There are so many differences between them that “Open Fully: can be overwhelming. This is “Open” for F01-F99 –
But this is the beginning of “Open Fully” [web page continues for many entries] –

- F01-F99 (Mental and behavioural disorders)
  - F01-F09 (Organic, including symptomatic, mental disorders)
    - F01 (Vascular dementia)
      - F01.0 (Vascular dementia of acute onset)
      - F01.1 (Multi-infarct dementia)
      - F01.2 (Subcortical vascular dementia)
      - F01.3 (Mixed cortical and subcortical vascular dementia)
      - F01.8 (Other vascular dementia)
  - F01.9 (Vascular dementia, unspecified)
- F03 (Unspecified dementia)

And here is the ending of “Open Fully” after many entries –

- F98.3 (Pica of infancy and childhood)
- F98.4 (Stereotyped movement disorders)
- F98.5 (Stuttering [stammering])
- F98.6 (Cluttering)
- F98.8 (Other specified behavioural and emotional disorders with onset usually occurring in childhood)
- F98.9 (Unspecified behavioural and emotional disorders with onset usually occurring in childhood)
- F99-F99 (Unspecified mental disorder)
- F99 (Mental disorder, not otherwise specified)

This is why I recommend not using “Open Fully”.

* You need to open and read the “Help” in “Display Options” to understand these options.
* Anywhere you see the Send button you can click it to start a search.
* Here is the entire “Help” file; print it; read it; understand it – https://wonder.cdc.gov/wonder/help/ucd.html
This is where you find intel about methodology, definitions, and links to more intel about CDC-Wonder. One item to point out is the information about suppression of data; this is the entire file –

What are the Assurance of Confidentiality constraints for the data?
Data reports for years 1989 and later must meet the NCHS data use restrictions. Vital statistics data are suppressed due to confidentiality constraints, in order to protect personal privacy. The term "Suppressed" replaces sub-national death counts, births counts, death rates and associated confidence intervals and standard errors, as well as corresponding population figures, when the figure represents zero to nine (0-9) persons.

As of December 12, 2011, additional privacy constraints apply to infant mortality statistics representing infant age groups and live births as the denominator population. When an infant mortality measure represents fewer than ten (0-9) infant deaths, all corresponding live birth population denominator figures are suppressed. When the infant mortality measure represents ten to nineteen (10-19) infant deaths, the number of deaths and live births are shown, but rates and associated measures are not shown. Race and Hispanic origin data are not available in this online database for infant age groups. However, race and Hispanic origin data are available for persons under one year of age in the other age groups, which use population estimates as population denominator data. Race and Hispanic origin detail for infant mortality statistics are available in the Linked Birth / Infant Death Records data collections.

Prior to May 23, 2011, data cells in tables for year 1989 and later years were suppressed only for single county-level data, when the data represented five or fewer (1-5) deaths for a time period less than three years, and the county’s total population in the April 1st, 2000 Census was fewer than one hundred thousand (100,000) persons. Prior to December 12, 2011, the same constraints applied to infant mortality statistics and all-ages mortality statistics.

Totals and sub-totals are suppressed when the value falls within scope of the suppression criteria, or when the summary value includes a single suppressed figure, in order to prevent the inadvertent disclosure of suppressed values.

The confidentiality constraints and use of the "Unreliable" flag are established by the original data providers. For more information, please contact the data providers.

One thing to take away from the above is that if you want data at the geo level of the entire nation, all races, all age groups, all everything, it’s not such a problem, but if you want data for small geographies and sub-populations, you’ve got problems.

* Remember to click the “Reset” button on the upper-right screen of the Request, Results, Map, or Chart web page to begin a new search. It’s important to completely clear your prior search.

We need to examine these two data display choices that you been seeing beside Request Form: MAP and CHART –

| Request Form | Results | Map | Chart |

For both Map and Chart do your basic search first, then choose “Map”
or “Chart”. The following is a simple search in OTIS (Online Tuberculosis Information System) –

* State
* 2012 – 2016 (column title reads 1993 – 2016)
* Alabama, Florida, Georgia, Mississippi, North and South Carolina
* Age group 45 – 64 years
* Revised occupation – Health Care
* Bridged race/ethnicity – All values
* Step 4 – accept all defaults
* Step 5 – accept all defaults except “Previous TB – No” and “Verification Criteria – Positive culture and Clinical case definition”
* Step 6 and 7 – accept all defaults

![OTIS TB Data 1993-2016 Results](image-url)
Now click the top-screen *Map* tab for this –
Then click the top *Chart* tab for this –

Steve Beleu, Oklahoma Dept. of Libraries
May 10, 2018