

Accessing SDWIS/FED drinking water data in Excel PivotTables

1	What's a PivotTable®?	1
	A few terms.....	1
2	What's in the 4 main PivotTables®?	3
2.1	Inventory.....	3
2.2	Violation type.....	3
2.3	Contaminant type.....	4
2.4	GPRA.....	4
3	How to use the PivotTables®	7
3.1	Just the basics.....	7
3.1.1	Adding or removing dimensions and measures.....	7
3.1.2	Filtering data members.....	8
3.1.3	Rearranging the dimensions and measures to facilitate analysis.....	9
3.1.4	Drilling down.....	9
3.1.5	Saving PivotTables®.....	10
3.2	Organizing data, printing, graphing.....	11
3.2.1	Adding and removing dimension subtotals.....	11
3.2.2	Adding or removing row or column totals.....	11
3.2.3	Changing labels.....	12
3.2.4	Sorting data in ascending or descending order.....	12
3.2.5	Organizing data using the Page area.....	12
3.2.6	Printing.....	13
	Graphing.....	14
3.3	Some additional functions not covered here.....	15
4	So, what do we make of all these numbers?	16
4.1	General tips and mistakes to avoid.....	16
4.1.1	Don't forget to select the 'Sum of...' measures each time you insert them into a PivotTable®.....	16
4.1.2	Don't refresh your data unless you have the original data tables on your computer... ..	16
4.1.3	Don't put measures in a Column, Row or Page area and don't put dimensions in the Data area.....	16
4.1.4	Don't forget about filtered or "hidden" data members.....	16
4.1.5	Don't add the number of systems in violation or population affected across violation types or contaminant types.....	16
4.1.6	Be careful when using population served data, especially when looking for trends	17
4.1.7	Organizing 'state' and 'source' data.....	17
4.2	Tips by PivotTable®.....	17
4.2.1	Inventory PivotTable®.....	17
4.2.2	Violation type PivotTable®.....	17
4.2.3	Contaminant type PivotTable®.....	18
4.2.4	GPRA PivotTable®.....	18
	Getting behind the numbers in the PivotTables®	19
4.3	Frozen databases used.....	19
4.4	Violation and contaminant codes.....	19
5	Contact information	19

Accessing SDWIS/FED drinking water data in MS Excel PivotTables®

These instructions will get you started using MS Excel PivotTables® that contain EPA Safe Drinking Water Information System (SDWIS/FED) data. No prior knowledge of spreadsheets or of MS Excel® is required. The following PivotTables® are described in these instructions:

1. Inventory—contains summary inventory data on water systems
2. Violation type—contains violations data by violation type
3. Contaminant type—contains violations data by contaminant types and rules
4. GPRA—contains data on health-based violations, and can be used to calculate GPRA (Government Performance and Results Act) data

These instructions are organized as follows: Section 1 describes what PivotTables® are and defines some terms. Section 2 lists what data are contained in the 4 main PivotTables®, and Section 3 tells you how to use them. Some mistakes to avoid are listed in Section 4.

And to make things even easier, each PivotTable® contains a full description, brief instructions, and tips/mistakes to avoid.

In addition, several additional PivotTables® have been built. You can download PivotTables® from:

www.epa.gov/safewater/data/getdata.html

1 What's a PivotTable®?

A PivotTable® is a Microsoft Excel® tool. To stoop to “techie” terms, they are multidimensional databases (MDBs) that provide online analytical processing, or OLAP. This means they enable you to quickly summarize, cross-tabulate, and analyze large amounts of data. You can pivot, or rotate, rows and columns to see different summaries of the source data, filter the data, and drill-down to the details in the underlying source data.

For example, you can easily find the number of systems reporting a certain type of violation, in a certain year, for a certain state. To see the data in context, you can compare the results to those of other states, or to other states in that region. To look for trends you can build a graph to compare the results across several years. To investigate an unusually high or low result you can see if the problem stems from a certain contaminant or rule, within a certain system type, size category, or primary source. Using the PivotTables® described in these instructions, you could probably perform this analysis in less than 15 minutes.

In addition, you can easily build PivotTables® to organize and analyze additional data, and to greatly simplify your SDWIS/FED queries.

A few terms

PivotTables® are primarily made up of “Dimensions” and “Measures.”

Measures contain the core information or source data—the numbers, the facts. These include the number of systems and population served by them, as well as the number of violations reported, the number of systems reporting violations, and the population served by these systems.

Dimensions organize or categorize the measures. Dimensions include state, water system type, size category, fiscal year, etc.

Each dimension contains various “**Members**.” The Size category dimension, for example, includes 5 members—one for each size category. The State dimension includes individual states as members.

A sample PivotTable® is shown below. It contains inventory data on water systems.

Type	gw/sw	Data	Size category				
			1 Very Small	2 Small	3 Medium	4 Large	5 Very Large
CWS	gw	# systems	28,829	10,414	2,512	1,372	68
		population served	4,536,976	14,122,773	14,357,264	34,471,159	18,901,262
	sw	# systems	3,075	3,626	1,844	1,904	279
		population served	632,797	5,654,251	10,997,486	56,575,146	104,746,819
NTNCWS	gw	# systems	16,641	2,608	69	13	1
		population served	2,313,342	2,521,825	335,447	242,366	110,000
	sw	# systems	546	180	17	6	1
		population served	88,277	215,917	83,282	244,204	169,846
TNCWS	gw	# systems	89,008	2,566	140	84	10
		population served	7,341,785	2,533,079	804,666	2,773,299	2,527,503
	sw	# systems	1,728	155	27	11	
		population served	171,175	187,437	152,289	340,296	

Here the measures are ‘# systems’ and ‘population served’ by them. Dimensions including ‘Size category’, ‘Type’, and ‘gw/sw’ categorize these measures. The members in each dimension are also listed. For example, ‘CWS’, ‘NTNCWS’, and ‘TNCWS’ are members in the ‘Type’ dimension.

2 What's in the 4 main PivotTables®?

Below is an overview of the types of data contained in each of the 4 main PivotTables®. This is followed by a detailed list of all the dimensions, their associated members, and the measures contained in each of the 4 main PivotTables®.

These tables are updated annually. Details are provided in Section 4.3.

2.1 Inventory

The Inventory PivotTable® contains summary statistical information on the number of water systems and their population served. Only active, current systems are included (some water systems are seasonal and may be inactive at the time the database is frozen).

The following measures—

- # of water systems and/or the
- population served by water systems

—can be categorized by any combination of the following dimensions:

- County
- State
- Region
- Water system type (this is explained in the table below)
- Primary source
- Size category
- Owner type (this is explained in the table below)

2.2 Violation type

The Violation type PivotTable® contains reported violations by violation type for fiscal years 1993-2000. Violation types include MCL, TT, M/R, Other, and Total (these acronyms are spelled-out in the table below).

The following measures—

- # of violations reported and/or the
- # of water systems reporting violations and/or the
- population served by water systems reporting violations

—can be categorized by any combination of the following dimensions:

- Fiscal year
- Violation type
- State
- Region
- Water system type
- Primary source
- Size category

2.3 Contaminant type

The Contaminant type PivotTable® contains reported violations for certain contaminants and rules for fiscal years 1993-2000. These include TCR/T, TTHM and other VOCs, SOCs, Nitrates and other IOCs, Radionuclides, LCR and SWTR (these acronyms are spelled-out in the table below). These are further organized by violation type: MCL, TT and M/R.

The following measures—

- # of violations reported and/or the
- # of water systems reporting violations and/or the
- population served by water systems reporting violations

—can be categorized by any combination of the following dimensions:

- Contaminant type or rule
- Violation type
- Fiscal year
- State
- Region
- Water system type
- Primary source
- Size category

2.4 GPRA

The GPRA PivotTable® contains data on systems reporting health-based violations (MCL and TT violations) for fiscal years 1993-2000. The ‘GPRA’ Dimension lists the percent of the total population served by systems reporting no health-based violations (remember that the GPRA goals apply only to CWSs).

The following measures—

- # of health-based violations reported and/or the
- # of water systems reporting health-based violations and/or the
- total # water systems
- % of water systems reporting health-based violations and/or the
- population served by water systems reporting health-based violations and/or the
- total population served by water systems
- % of total population served by water systems reporting health-based violations and/or the
- % of total population served by water systems not reporting any health-based violations (GPRA)

—can be categorized by any combination of the following dimensions:

- Fiscal year
- State
- Region
- Water system type
- Primary source
- Size category

Dimensions included in all 4 PivotTables

Dimension	Members	
State	Numeric states (01 to 10) refer to tribal systems in that respective region.	
Region	I	CT, ME, MA, NH, RI, VT
	II	NJ, NY, PR (Puerto Rico), VI (Virgin Islands)
	III	DE, DC, MD, PA, VA, WV
	IV	FL, FL, GA, KY, MS, NC, SC, TN
	V	IL, IN, MI, MN, OH, WI
	VI	AR, LA, NM, OK, TX
	VII	IA, KS, MO, NE
	VIII	CO, MT, ND, SD, UT, WY
	IX	AZ, CA, HI, NV, AS (American Samoa), GU (Guam), MP (Northern Marianas Islands), PW (Palau)
	X	AK, ID, OR, WA
Type	CWS	Community Water System
	NTNCWS	Non-Transient Non-Community Water System
	TNCWS	Transient Non-Community Water System
Source	Primary source of water:	
	SW	surface water
	SWP	purchased surface water
	GU	ground water under the influence of surface water
	GUP	purchased ground water under the influence of surface water
	GW	ground water
Gw/sw	SW	includes SW, SWP, GU, GUP
	GW	includes GW and GWP
Size category	Based on average daily population served:	
	1 Very small	25-500
	2 Small	501-3,300
	3 Medium	3,301-10,000
	4 Large	10,001-100,000
5 Very large	>100,000	

Dimensions included in specific PivotTables

Pivot table:	Inventory				
Dimension	Violation type	Contaminants		GPRA	Members
County	X				County served
Owner	X				F Federal government
					S State government
					L Local government
					M Mixed public/private
					N Native American
					P Private
FY	X	X	X		Violation type and Contams tables go back to 1993. GPRA goes back to 1990. A fiscal year (FY) runs from October 1 st to September 30 th .
Vtype	X (1-5)	X (1-3)			1 MCL Maximum Contaminant Level violation
					2 TT Treatment Technique violation
					3 M/R Monitoring or reporting violation
					4 Other Other violation, including public notice violation
					5 Total All violations
Ctype		X			1 TCR/T Total Coliform Rule and Turbidity
				2 TTHM Total Trihalomethanes, a Volatile Organic Chemical (VOC)	
				3 VOC Other Volatile Organic Chemicals	
				4 SOC Synthetic Organic Chemicals	
				5 NIT Nitrates	
				6 IOC Other Inorganic Chemicals	
				7 RAD Radionuclides	
				8 LCR Lead and Copper Rule	
				9 SWTR Surface Water Treatment Rule	

Measures included in specific PivotTables

Pivot table: Inventory

Field	Violation type			GPRA	Items
	Inventory	Contaminant type			
# systems	X				Number of active, current systems
Pop	X				Population served by active, current systems
# viols		X	X	X	# of violations reported
# in viol		X	X	X	# of systems reporting violations
Pop		X	X	X	Population served by systems reporting violations
Tot # sys				X	Total # systems (same as '# systems' above)
% sys in viol				X	% systems reporting health-based violations
Total pop				X	Total population served (same as 'pop' above)
% pop in viol				X	% population served by systems reporting health-based violations
GPRA				X	% of population served by systems not reporting any health-based violations (GPRA applies to CWSs only)

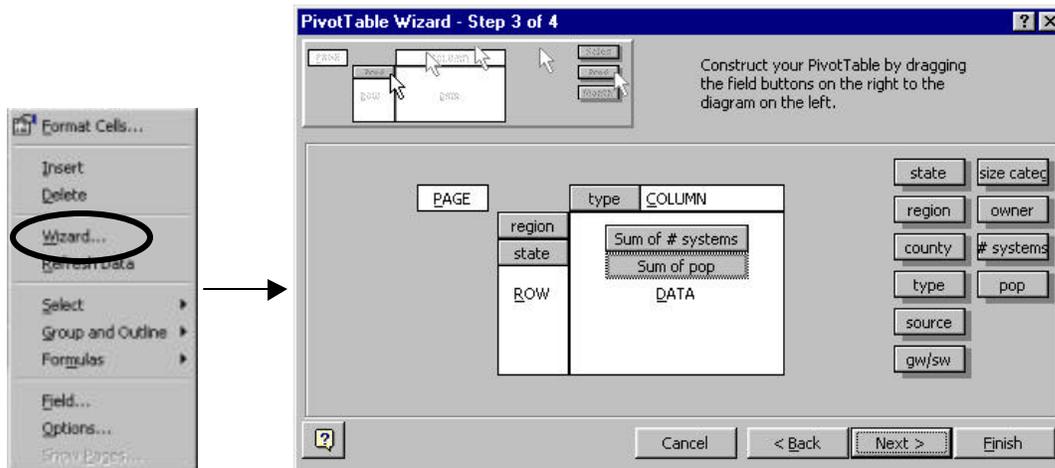
3 How to use the PivotTables®

This section covers basic information to get you started using PivotTables®. More information is available in the help menus in Excel®. Also, feel free to give me a call if you have any questions.

3.1 Just the basics

3.1.1 Adding or removing dimensions and measures

Use the Wizard to add or remove dimensions and measures in a PivotTable®. To get to the Wizard, select any cell inside the PivotTable® and right-click on it. Then select ‘Wizard...’.



The Wizard lists all of the dimensions and measures available in the PivotTable® (which in this case is the Inventory PivotTable®) on the right side. The measures (in this case “# systems” and “pop”) are listed last, by convention.

The dimensions and measures placed in the “Row area” “Column area” and “Data area” on the left side of the Wizard reflect the structure of the PivotTable®. The Wizard shown above underlies the following PivotTable®:

	A	B	C	D	E	F	G
1				Type			
2	Region	State	Data	CWS	NTNCWS	TNCWS	Grand Total
3		1	1	Sum of # systems	1	2	3
4				Sum of pop	120	60	670
5		CT		Sum of # systems	597	590	3,437
6				Sum of pop	2,644,154	115,348	2,976,839
7		MA		Sum of # systems	521	229	837
8				Sum of pop	6,990,460	75,318	230,704
9		ME		Sum of # systems	399	376	1,302
10				Sum of pop	609,447	72,471	208,368
11		NH		Sum of # systems	676	450	1,040

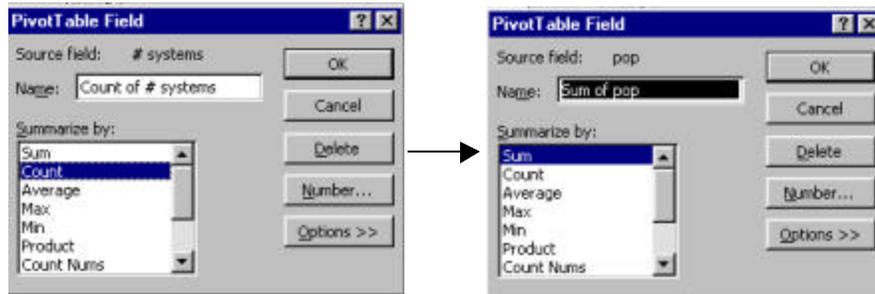
As reflected in the Wizard:

- ‘Type’ is selected in the Column area
- ‘region’ and ‘state’ are selected in the Row area
- both measures—“# systems” and “pop”—are selected in the Data area

You can construct or alter a PivotTable® by dragging dimensions and measures to Row, Column, or Data areas. Make sure you don’t place dimensions fields in the Data area, or measures in the Row, Column, or Page areas.

Placing a dimension in the Page area will enable you to list the data one member at a time. For example, if you placed 'Region' in the Page area you could look at the data for any single region at a time. With this feature, you also have the option to look at the data for all regions at once. In which case nothing is filtered, and the Page area can serve as a place to store unused dimensions.

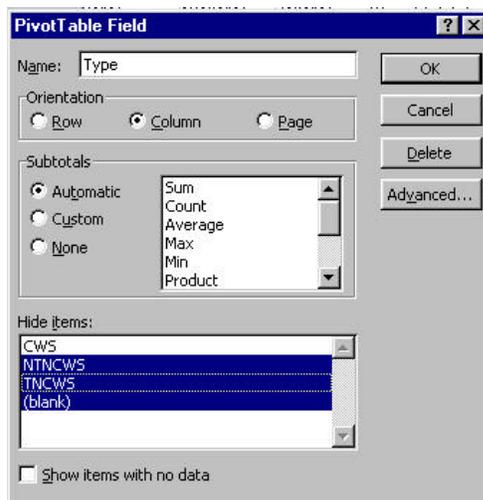
It's important to remember that whenever you drag a measure onto the Data area, you'll need to specify that the 'Sum' of the values is used. The default is 'Count'. To correct this, double-click on the measure; the following screen will appear:



Select 'Sum'.

3.1.2 Filtering data members

You can filter-out members within any dimension that's currently used in a PivotTable®. Select a dimension (which are colored gray) and either right-click or double-click on it. The screen below will appear. Then select which members you'd like to hide:



In the example above, we've hidden NTNCWSs and TNCWSs (and blanks) from the 'Type' dimension. Only CWSs will be included.

Be careful: it's easy to forget you've hidden members. It's a good idea to frequently right-click on a dimension and double-check if any members are hidden. Also be aware that if you move a dimension out of the PivotTable® area and bring it back in, any members you'd previously hidden will remain hidden. However, the data from all members will be counted if the dimension is removed from your PivotTable®.

3.1.3 Rearranging the dimensions and measures to facilitate analysis

You can rearrange dimensions and measures in a PivotTable® “on the fly” without using the Wizard. To do this, point to a dimension you’re interested in moving, press and hold the mouse button, and move it around. You can move it from the Row area to the Column area or the Page area. You can place it before or after another dimension in the Row or Column area, or move it to or from the Page area. In addition, you can remove a dimension from your PivotTable® by simply dragging it out of the PivotTable® area. Fiddle around with this for a while to see how it works.

I usually build a PivotTable® by selecting all the required dimensions and measures I’ll need; then I organize it “on the fly” by dragging them around. Below is a simple example.

		type		Data			
		CWS		NTNCWS		TNCWS	
FY	vtype	Sum of # in viol	Sum of pop	Sum of # in viol	Sum of pop	Sum of # in viol	Sum of pop
1999	1 MCL	3,321	11,079,343	958	292,367	3,876	646,469
	2 TT	1,057	15,854,722	178	53,500	140	42,582
	4 Other	627	2,529,718	88	34,947	498	63,264
1998	1 MCL	3,746	10,393,015	945	232,755	3,799	481,642
	2 TT	1,105	18,485,545	103	23,963	90	58,801
	3 M/R	10,002	28,313,863	3,942	994,786	15,997	2,102,061
	4 Other	899	2,369,301	157	47,772	1,115	143,464
	5 Total	13,024	48,945,871	4,672	1,187,783	18,771	2,458,001

In the (Violation type) PivotTable® above, the measures are listed in a columnar orientation. The default is a row orientation. In the second PivotTable®, I moved the ‘type’ field from the Column area to the Row area and placed before ‘FY’. In the last PivotTable®, I moved ‘vtype’ before the ‘type’ and ‘FY’ dimensions.

		Data	
type	FY	vtype	Sum of # in viol
CWS	1999	1 MCL	3,321
		2 TT	1,057
		4 Other	627
	1998	1 MCL	3,746
		2 TT	1,105
NTNCWS	1999	1 MCL	958
		2 TT	178
		4 Other	88
	1998	1 MCL	945
		2 TT	103
TNCWS	1999	1 MCL	3,876
		2 TT	140
		4 Other	498
	1998	1 MCL	3,799
		2 TT	90

		Data	
vtype	type	FY	Sum of # in viol
1 MCL	CWS	1999	3,321
		1998	3,746
		NTNCWS	1999
	TNCWS	1998	945
		1999	3,876
1 MCL Total			16,645
2 TT	CWS	1999	1,057
		1998	1,105
		NTNCWS	1999
	TNCWS	1998	103
		1999	140
2 TT Total			2,673
3 M/R	CWS	1998	10,002
	NTNCWS	1998	3,942
	TNCWS	1998	15,997
3 M/R Total			29,941
4 Other	CWS	1999	627
		1998	899
	NTNCWS	1999	88
		1998	157
		TNCWS	1999
4 Other Total			1,115

In short, these data can be displayed in any combination.

3.1.4 Drilling down

You can get to the core data in any PivotTable® data cell by double-clicking on the data in any data cell. Excel® will automatically create another worksheet containing the core data.

As an example of how to drill down, suppose you were interested in finding out more about CWSs served primarily by GU (ground water under the influence of surface water) in Region I. A PivotTable® which contains the required dimensions and measures is shown below.

		Type	Source						
Region	Data	CWS						CWS Total	
		GU	GUP	GW	GWP	SW	SWP		
1	Sum of # systems	13		2,206	29	296	173		2,717
	Sum of pop	24,727		2,678,006	19,361	6,528,771	3,197,735		12,448,600
2	Sum of # systems	3		2,655	60	666	201		3,585
	Sum of pop	2,747		8,149,015	89,373	19,303,964	1,888,324		29,433,443
3	Sum of # systems	137		3,749	93	591	496		5,066
	Sum of pop	196,221		2,896,467	138,532	16,224,237	4,017,358		23,472,815
4	Sum of # systems	63	21	7,373	404	696	1,081		9,638
	Sum of pop	234,649	34,320	21,901,895	922,630	18,521,572	5,470,516		47,085,582
5	Sum of # systems	23		6,125	341	407	907		7,803
	Sum of pop	511,633		13,114,663	737,925	16,445,245	8,593,270		39,402,736
6	Sum of # systems	18	1	6,026	379	704	1,251		8,379
	Sum of pop	86,115	339	11,660,043	501,081	15,247,537	5,162,282		32,657,397

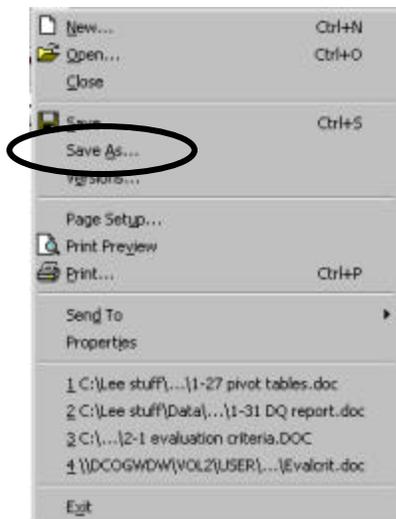
When you double-click on a data cell (in this example, the data cell that contains 13 systems), Excel® will create the following worksheet:

	A	B	C	D	E	F	G	H	I
1	State	Region	Source	gw/sw	Type	Size category	Owner	# systems	pop
2	CT	1	GU	sw	CWS	1 Very Small	?	1	50
3	CT	1	GU	sw	CWS	2 Small	?	1	584
4	CT	1	GU	sw	CWS	3 Medium	Private	1	3,500
5	CT	1	GU	sw	CWS	1 Very Small	Private	2	290
6	VT	1	GU	sw	CWS	1 Very Small	?	3	455
7	VT	1	GU	sw	CWS	2 Small	?	2	3,480
8	VT	1	GU	sw	CWS	4 Large	?	1	10,420
9	CT	1	GU	sw	CWS	3 Medium	Local	1	5,918
10	ME	1	GU	sw	CWS	1 Very Small	Fed	1	30

Unfortunately, the core data underlying these 4 main PivotTables® doesn't go all the way down to the water system level due to Excel®'s 65,000 row limitation. Because of this, summary data is provided at the core level.

3.1.5 Saving PivotTables®

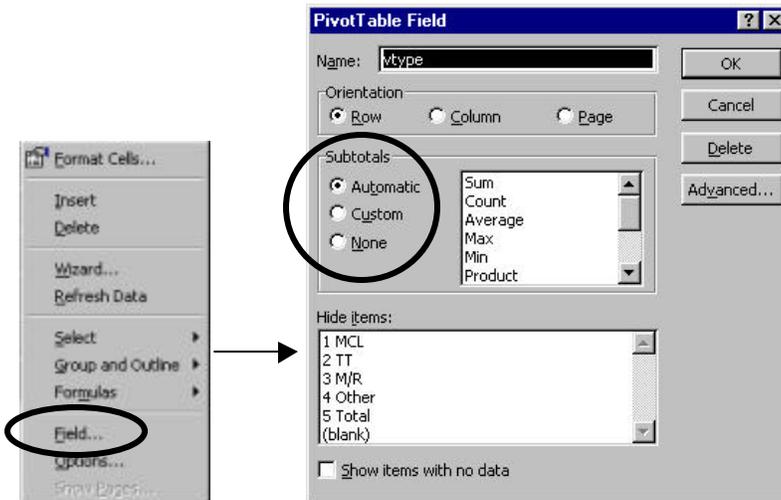
The PivotTables® are read-only, which means they will appear in the same format each time that you open them. If you want to save changes to a PivotTable® you've modified, just save it under a different name. To do this, press the 'File' menu and select 'Save As...' and save it under a different name.



3.2 Organizing data, printing, graphing

3.2.1 Adding and removing dimension subtotals

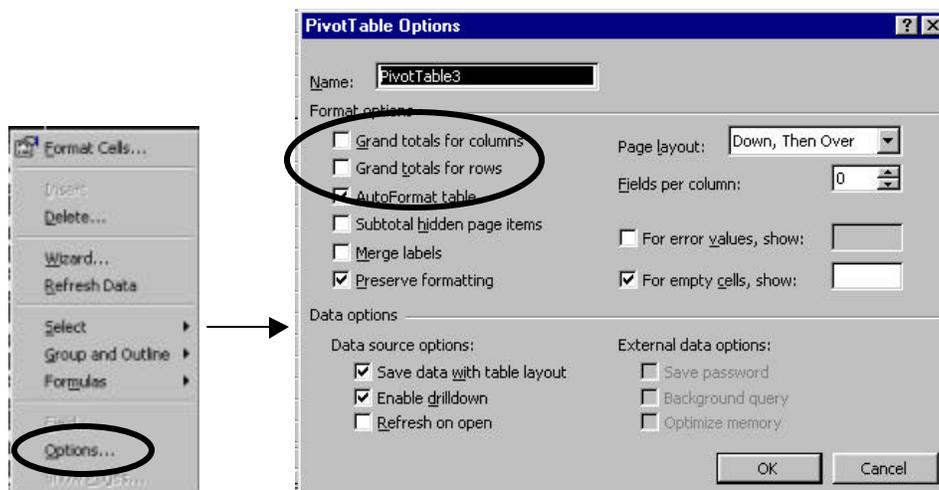
Subtotals of data items in a dimension can be very helpful. To add or remove them, first point your cursor on the dimension of interest. Next, right-click your mouse, and then select 'Field...', as shown below:



Be careful when using subtotals and row or column totals. It's easy to sum things that shouldn't be summed, such as across violation types or contaminants/rules. For example, if you simply added the number of systems having MCL, M/R, and Other violation types, one system having one or more of each type would be counted three times instead of once. In the example above, you should select 'None' as the subtotal of the 'vtype' dimension.

3.2.2 Adding or removing row or column totals

To add or remove row or column totals right-click on any data cell in the PivotTable®, and then select 'Options...'. This is shown below:



You can also remove row or column totals by selecting them in the PivotTable®, right-clicking your mouse, and selecting 'Delete...'

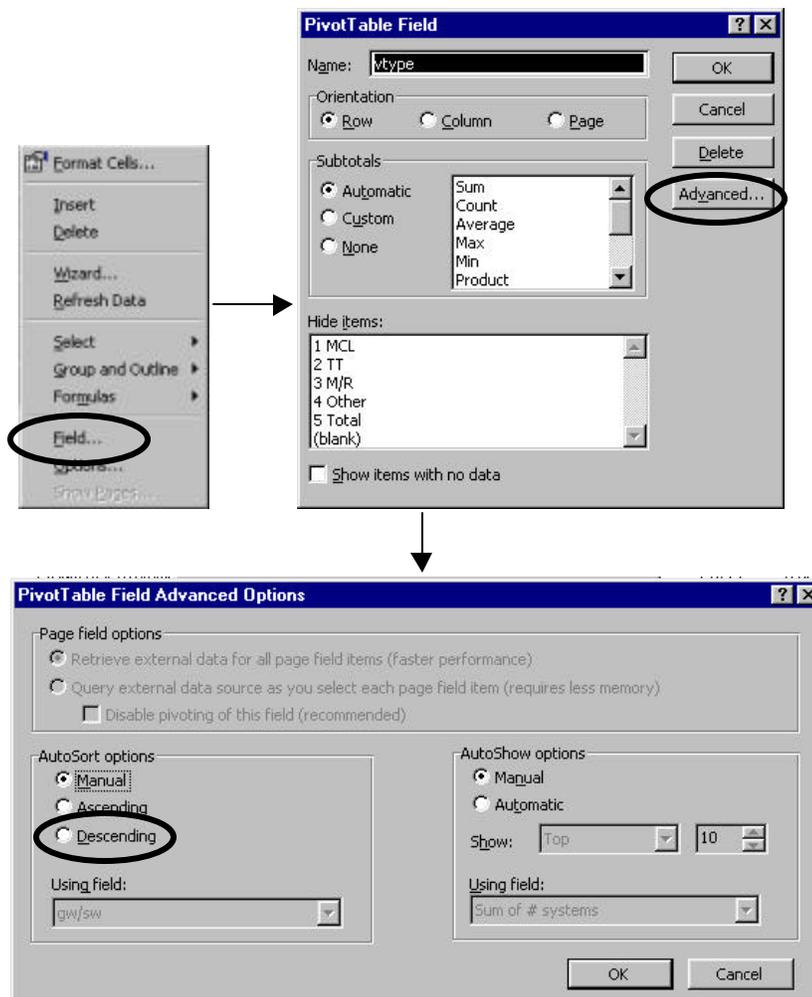
As mentioned above, be careful when using totals by row or column. Summing across violation types or contaminants/rules can yield erroneous results.

3.2.3 Changing labels

To make your PivotTable® labels more understandable, you can simply click on the label and type-over them. For example, instead of listing State “01” you might want to type something like “Tribal systems in Region 1.” These changes will be lost if you remove the field from the PivotTable® and later restore it.

3.2.4 Sorting data in ascending or descending order

You can choose whether members in Row/Column fields are listed in ascending or descending order. For example, when listing FY in the Row area, my personal preference is to list the most recent year first. To sort dimension members in descending order double-click on the field, select ‘Advanced...’ and then ‘Descending’ under ‘AutoSort options’.



3.2.5 Organizing data using the Page area

As discussed above, this feature enables you to quickly filter-out (hide) all but a single member in a dimension. For example, if you’re primarily interested in data for a certain state, simply place

'State' in the Page area and select that individual state from the drop-down toolbar. This is quicker and easier than placing 'State' in the Row or Column area, right-clicking your mouse on the 'State' dimension, selecting 'Field...,' and hiding all the other states one at a time.

The Page area can also help you organize large amounts of data. With a dimension in the Page area, you can make a separate PivotTable® for each member in that dimension. For example, with the 'Region' dimension listed in the Page area you can make a separate PivotTable® in a separate worksheet for each region. To do this, point to a cell in the PivotTable®, and right-click your mouse. The following screen will appear.



Select 'Show Pages' to make a PivotTable® for each member in the dimension.

3.2.6 Printing

It can be very easy to make large or even huge PivotTables®, and so difficult to organize them to print in a logical way. Here are a few quick tips.

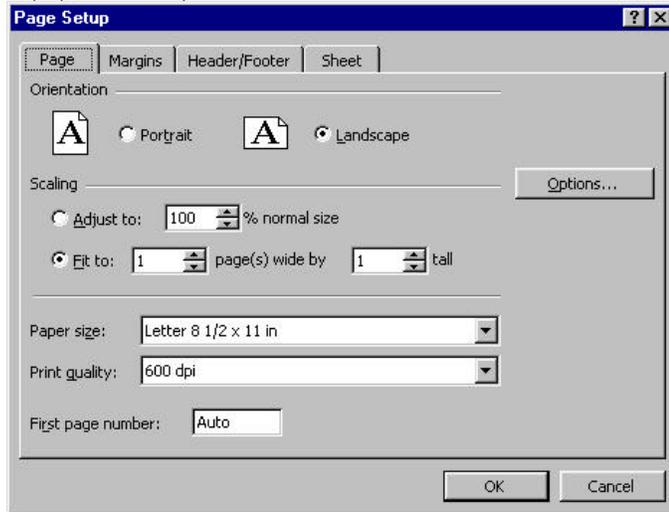
Before printing, press the 'Print Preview' menu button. The button looks like a white page with a magnifying glass:



You can also get there from the main menu bar by selecting 'File' and then 'Print Preview'. It will show you how your PivotTable® will set up on the page, as well as how many pages it will take to print. The top menu bar is shown below:



From here press the "Setup..." menu button to alter how the spreadsheet will print. The 'Page Setup' screen will appear:



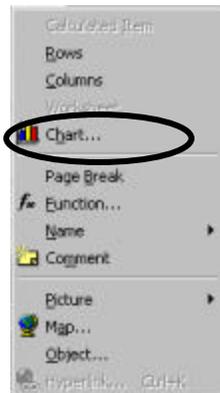
You can see whether your PivotTable® will print better in either Landscape or Portrait orientation. To fit the most on a page, select Landscape and then Portrait and see how small the data will print. Try not go below 60% or it will be difficult to read.

For large PivotTables® that will take several pages to print, the 'Page Break Preview' feature will enable you to select where each page break will occur. You can get there by pressing the 'Print Preview' button and then the 'Page Break Preview' button. It's also usually a good idea to print row and/or column headings at the top of each page. You can specify these by selecting the 'Sheet' button from the 'Page Setup' screen.

Graphing

Excel® has powerful graphing capabilities. There are 'Wizards' to help you make graphs of your data, and a little experimenting will get you far. Here's a few quick tips to get you started:

- Select only one measure at a time
- Remove all subtotals and totals.
- Select the entire PivotTable® with your mouse by clicking in the upper left hand cell of the PivotTable®. You'll know the entire PivotTable® is selected if the entire field is highlighted.
- Under the 'Insert' menu, select 'Chart...'



From there, a 'Wizard' will guide you. Once you've made a chart you can change the underlying data and it will graph automatically. For example, suppose you make a chart using the Violation type PivotTable® and graph the number of violations by FY (remember, you've got to include

the 'Vtype' field when using this PivotTable®). You can replace the '# viols' measure with '# in viol' or 'pop' measures and Excel® will automatically change the chart. Likewise when printing one member at a time in any dimension.

3.3 Some additional functions not covered here

There are numerous elements of PivotTables® not covered here. If you would like to learn more about them, here are a couple of functions you might want to look into.

One of them is the PivotTable® toolbar. Using it will enable you to go directly to the Wizard or to format a dimension or measure without right-clicking. There are additional functions there as well. I haven't discussed how to use the toolbar here since some of the functions are slightly more advanced. However, fiddling will get you far: there are excellent help menus available, and Excel® is very well thought-out—learning how to do additional things is almost intuitive.

Another powerful tool is the ability of Excel® to calculate measures. Right-click on a cell in the PivotTable® and select 'formulas', then 'Calculated field...' to get started. The GPRA PivotTable® contains some simple calculated measures.

4 So, what do we make of all these numbers?

Now you can easily extract and analyze large amounts of data from SDWIS/FED. Unfortunately, there's still a few ways to get incorrect or meaningless numbers. Please spend some time to understand this section.

4.1 General tips and mistakes to avoid

4.1.1 Don't forget to select the 'Sum of...' measures each time you insert them into a PivotTable®

The default is 'Count of...'. To correct this, double-click the measure, and select "Sum." See 3.1.1.

4.1.2 Don't refresh your data unless you have the original data tables on your computer

The core data tables are not included with the 4 main PivotTables®. If you select a cell in the PivotTable® and right-click your mouse, you'll see the 'Refresh Data' button. If you select this and you don't have the core data tables on your computer, the program will crash and you'll have to reboot.

4.1.3 Don't put measures in a Column, Row or Page area and don't put dimensions in the Data area

You'll end up with meaningless results.

4.1.4 Don't forget about filtered or "hidden" data members

As discussed above, it's easy to forget you've hidden data members from a dimension. It's especially easy when dimensions are placed in the Page area. Leaving them out will lower your overall data totals.

4.1.5 Don't add the number of systems in violation or population affected across violation types or contaminant types.

Since systems can have many different violation types and/or contaminant/rule violations, the number of systems in violation (and population served) cannot simply be added. For example, if you added the number of systems having MCL, M/R, and Other violation types, one system having one or more of each type would be counted three times instead of once.

There are ways to get accurate counts of systems and population affected across violation types:

- The GPRA PivotTable® can list the number of systems reporting health-based violations (MCL and/or TT violations), and the population affected.
- The Violation type PivotTable® contains the field "5 Total," that lists the number of systems, and population served, by water systems reporting any type of violation.

4.1.6 Be careful when using population served data, especially when looking for trends

When analyzing population served by systems having a violation, be careful of reading too much into the data, or to draw conclusions looking only at this parameter. Especially if there's only a small number of systems in violation, and/or those in violation tend to be smaller systems, one or a few very large systems can significantly skew the results. For example, there would appear to be a serious problem if the total population affected by a contaminant, which has been roughly 500,000 a year, suddenly jumps to 10,000,000. This increase, by a factor of 20, may indeed indicate a serious health problem. However, it may also be due to one very large system having a violation, and is therefore not indicative of a trend.

When we developed the Trends Report we wanted to cite trends primarily using population affected. However, the findings were generally unreliable and could even be misleading given the fact that one or a few very large systems could significantly skew the findings, as discussed above. Many of the erratic jumps in population served estimates are due to what we call "noise" in the data.

The number of systems in violation provides a more reliable measure of compliance when assessing trends, as it is not susceptible to bias by large systems. Looking at the total number of violations can also be subject to bias by one or a few systems having several violations, but this measure is still significantly less volatile than the population served numbers.

The higher the population affected by systems in violation (which comes with more systems in violation), the less susceptible will be the results to a very large system, and the more confidently you can use the data. So be careful not to parse the data too far (by state, by water system type, etc.), as this lessens the total population served and increases the affect of very large systems on the results.

4.1.7 Organizing 'state' and 'source' data

When using the 'state' dimension, if you also specify the 'region' dimension, place it before the 'state' dimension, and subtotal it, you'll get results by state and region in the same PivotTable®. In the same way, you can place the 'gw/sw' dimension before 'source' and subtotal it.

4.2 *Tips by PivotTable®*

4.2.1 Inventory PivotTable®

- Since several states can have the same county names, don't use the 'County' dimension unless its preceded by the 'State' dimension.
- Don't add population served across system types. Since all these types of water systems can serve an individual there's a lot of double-counting in the totals. For this reason, OGWDW usually lists the total population served by PWSs as those served by CWSs only.

4.2.2 Violation type PivotTable®

- Don't make a PivotTable® without including the 'Vtype' dimension in the Row or Column area. If you do, the sum of MCL+TT+M/R+Other+Total will be listed, which will result in double the number of violations.
- Don't use subtotals or grand totals across 'Vtype'. These are already contained in the Item '5 Total'.

4.2.3 Contaminant type PivotTable®

- Don't make a PivotTable® without specifying the 'Ctype' dimension in the Row or Column area.
- Don't use subtotals or totals to sum the number of systems in violation or population affected across Ctype. You can use the 'Violation type' PivotTable® to sum by violation type, or overall. In addition, you can use the 'GPRA' PivotTable® if you want to get the total number of systems and population affected by systems reporting any MCL and/or TT violations.

4.2.4 GPRA PivotTable®

- For the reasons discussed in 4.1.6, it's not a good idea to parse the GPRA numbers too far. In my opinion, you can get more reliable estimates for analyzing trends using the number of systems reporting violations. Each system reporting a certain violation is counted equally, as opposed to estimates of the population affected, where a Very Large system can easily skew the results.
- In the PivotTables® GPRA estimates are included for all water system types. Remember, however, that the official GPRA measure only includes CWSs.

Getting behind the numbers in the PivotTables®

4.3 Frozen databases used

Each PivotTable® contains a list of which frozen databases were used to generate the data. They are as follows:

- 00Q4 for the inventory PivotTable®. This table was frozen in January 2001.
- 00Q4 for FY2000 violations data, except for Chemical M/R violations. These tables were frozen in January 2001.
- 01Q1 for FY2000 Chemical M/R violations. These tables were frozen in April 2001.
- 99Q4 for FY99 violations data, except for Chemical M/R violations. These tables were frozen in January 2000.
- 00Q1 for FY99 Chemical M/R violations. These tables were frozen in April 2000.
- 98Q4 for FY98 violations data, except for Chemical M/R violations. These tables were frozen in January 1999.
- 99Q1 for FY98 Chemical M/R violations. These tables were frozen in April 1999.
- 98Q1 for FY97 and earlier violations data. These tables were frozen in April 1998.

4.4 Violation and contaminant codes

The workbook containing the 4 main PivotTables® also includes a worksheet containing the contaminant codes and violation type codes that underlie them. Later we'll provide a description of the MS Access queries that used these codes to generate the data in the PivotTables®.

5 Contact information

PivotTables®, such as the 4 outlined in this paper, are one piece of a continuum of products being developed to make SDWIS/FED data more accessible and usable. Your feedback will help us improve these products and better assess your SDWIS/FED data needs. We can also make PivotTables® to meet your specific needs, and show you how to make your own.

Contact Lee Kyle to provide feedback, ask questions, or to learn more about some of the other PivotTables® available. (202) 260-1154 or kyle.lee@epa.gov.

You can download copies of the latest PivotTables® and instructions from the web at:

<http://www.epa.gov/safewater/data/getdata.html>