

## DATABASE MANAGERS

(Data for this tutorial at [www.peteraldhous.com/Data](http://www.peteraldhous.com/Data))

We've already seen how spreadsheets can filter data and calculate subtotals. Database managers, such as Access, can handle larger datasets, and with practice are more flexible and nimble for filtering, grouping and making calculations on subsets of data.

They also allow you to join multiple data tables into one, or match records across different datasets, if they have common fields – which can be a powerful tool. Again, we'll work with data used in reporting this story: <http://www.newscientist.com/article/dn18806-revealed-pfizers-payments-to-censured-doctors.html>.

## Basic database queries

Open **Pfizer.mdb** under the **File** tab, then double click to open the table **Pfizer**.

ID	org_indiv	first_plus	first	last	city	state	category	cash	other	total
1	3-D MEDICAL S	STEVEN BRUCE	STEVEN	DEITELZWEIG	NEW ORLEANS	LA	Professional A	\$2,625.00	\$0.00	\$2,625.00
2	AA DOCTORS, I	AAKASH MOHA	AAKASH	AHUJA	PASO ROBLES	CA	Expert-Led For	\$1,000.00	\$0.00	\$1,000.00
3	ABBO, LILIAN N	LILIAN MARGAI	LILIAN	ABBO	MIAMI	FL	Business Relat	\$0.00	\$448.00	\$448.00
4	ABBO, LILIAN N	LILIAN MARGAI	LILIAN	ABBO	MIAMI	FL	Meals	\$0.00	\$119.00	\$119.00
5	ABBO, LILIAN N	LILIAN MARGAI	LILIAN	ABBO	MIAMI	FL	Professional A	\$1,800.00	\$0.00	\$1,800.00
6	ABDULLAH RAF	ABDULLAH	ABDULLAH	RAFFEE	FLINT	MI	Expert-Led For	\$750.00	\$0.00	\$750.00
7	ABEBE, SHEILA	SHEILA Y	SHEILA	ABEBE	INDIANAPOLIS	IN	Educational Ite	\$0.00	\$47.00	\$47.00
8	ABEBE, SHEILA	SHEILA Y	SHEILA	ABEBE	INDIANAPOLIS	IN	Expert-Led For	\$825.00	\$0.00	\$825.00
9	ABILENE FAMIL	GALEN CHRIS	GALEN	ALBRITTON	ABILENE	TX	Professional A	\$3,000.00	\$0.00	\$3,000.00
10	ABOLNIK, IGOR	IGOR Z	IGOR	ABOLNIK	PROVO	UT	Business Relat	\$0.00	\$396.00	\$396.00
11	ABOLNIK, IGOR	IGOR Z	IGOR	ABOLNIK	PROVO	UT	Expert-Led For	\$1,750.00	\$0.00	\$1,750.00
12	ABOLNIK, IGOR	IGOR Z	IGOR	ABOLNIK	PROVO	UT	Meals	\$0.00	\$58.00	\$58.00
13	ABRAKSIA, SA	SAMIR	SAMIR	ABRAKSIA	BEACHWOOD	OH	Business Relat	\$0.00	\$88.00	\$88.00
14	ABRAKSIA, SA	SAMIR	SAMIR	ABRAKSIA	BEACHWOOD	OH	Expert-Led For	\$2,000.00	\$0.00	\$2,000.00
15	ABRAKSIA, SA	SAMIR	SAMIR	ABRAKSIA	BEACHWOOD	OH	Meals	\$0.00	\$189.00	\$189.00
16	ABRAKSIA, SA	SAMIR	SAMIR	ABRAKSIA	BEACHWOOD	OH	Professional A	\$2,500.00	\$0.00	\$2,500.00
17	ABRAMSON, S	STEVEN BARRY	STEVEN	ABRAMSON	NEW YORK	NY	Business Relat	\$0.00	\$38.00	\$38.00
18	ABRAMSON, S	STEVEN BARRY	STEVEN	ABRAMSON	NEW YORK	NY	Professional A	\$4,400.00	\$0.00	\$4,400.00
19	ABUZZAHAB, F	FARUK S	FARUK	ABUZZAHAB	MINNEAPOLIS	MN	Business Relat	\$0.00	\$2,074.00	\$2,074.00
20	ABUZZAHAB, F	FARUK S	FARUK	ABUZZAHAB	MINNEAPOLIS	MN	Meals	\$0.00	\$218.00	\$218.00
21	ABUZZAHAB, F	FARUK S	FARUK	ABUZZAHAB	MINNEAPOLIS	MN	Professional A	\$1,750.00	\$0.00	\$1,750.00
22	ABUZZAHAB, M	MARY JENNIFE	MARY	ABUZZAHAB	SAINT PAUL	MN	Business Relat	\$0.00	\$154.00	\$154.00
23	ABUZZAHAB, M	MARY JENNIFE	MARY	ABUZZAHAB	SAINT PAUL	MN	Expert-Led For	\$1,000.00	\$0.00	\$1,000.00
24	ACADIA WOMEN	MICHELLE MCD	MICHELLE	OWENS	CROWLEY	LA	Expert-Led For	\$4,000.00	\$0.00	\$4,000.00
25	ACCACHA, SIH	SIHAM DONIAZ	SIHAM	ACCACHA	MINEOLA	NY	Expert-Led For	\$1,250.00	\$0.00	\$1,250.00
26	ACCACHA, SIH	SIHAM DONIAZ	SIHAM	ACCACHA	MINEOLA	NY	Meals	\$0.00	\$93.00	\$93.00
27	ACEVEDO MAR	IRIS ARLENE	IRIS	ACEVEDO MAR	CAGUAS	PR	Expert-Led For	\$750.00	\$0.00	\$750.00
28	ACEVEDO MAR	IRIS ARLENE	IRIS	ACEVEDO MAR	CAGUAS	PR	Meals	\$0.00	\$59.00	\$59.00
29	ACKERMAN, I	IVAN FOSTER	IVAN	ACKERMAN	BRANDON	FL	Expert-Led For	\$1,250.00	\$0.00	\$1,250.00
30	ACOSTA, LUIS S	LUIS SILVIO	LUIS	ACOSTA	HOUSTON	TX	Expert-Led For	\$1,000.00	\$0.00	\$1,000.00
31	ADAM LANDSMAN	ADAM S	ADAM	LANDSMAN	BOSTON	MA	Professional A	\$3,000.00	\$0.00	\$3,000.00
32	ADAM ROSEN	ADAM MICHAEL	ADAM	ROSEN	CLEARWATER	FL	Business Relat	\$0.00	\$41.00	\$41.00
33	ADAM ROSEN	ADAM MICHAEL	ADAM	ROSEN	CLEARWATER	FL	Expert-Led For	\$2,400.00	\$0.00	\$2,400.00
34	ADAMS, SANDI	SANDRA GAIL	SANDRA	ADAMS	SAN ANTONIO	TX	Professional A	\$12,840.00	\$0.00	\$12,840.00
35	ADDONA, TOM	TOMMASO	TOMMASO	ADDONA	NEW YORK	NY	Business Relat	\$0.00	\$39.00	\$39.00
36	ADDONA, TOM	TOMMASO	TOMMASO	ADDONA	NEW YORK	NY	Expert-Led For	\$750.00	\$0.00	\$750.00
37	ADDONA, TOM	TOMMASO	TOMMASO	ADDONA	NEW YORK	NY	Meals	\$0.00	\$109.00	\$109.00
38	ADLER, DAVID	DAVID ELLIOTT	DAVID	ADLER	PORTLAND	OR	Business Relat	\$0.00	\$1,062.00	\$1,062.00
39	ADLER, DAVID	DAVID ELLIOTT	DAVID	ADLER	PORTLAND	OR	Meals	\$0.00	\$390.00	\$390.00

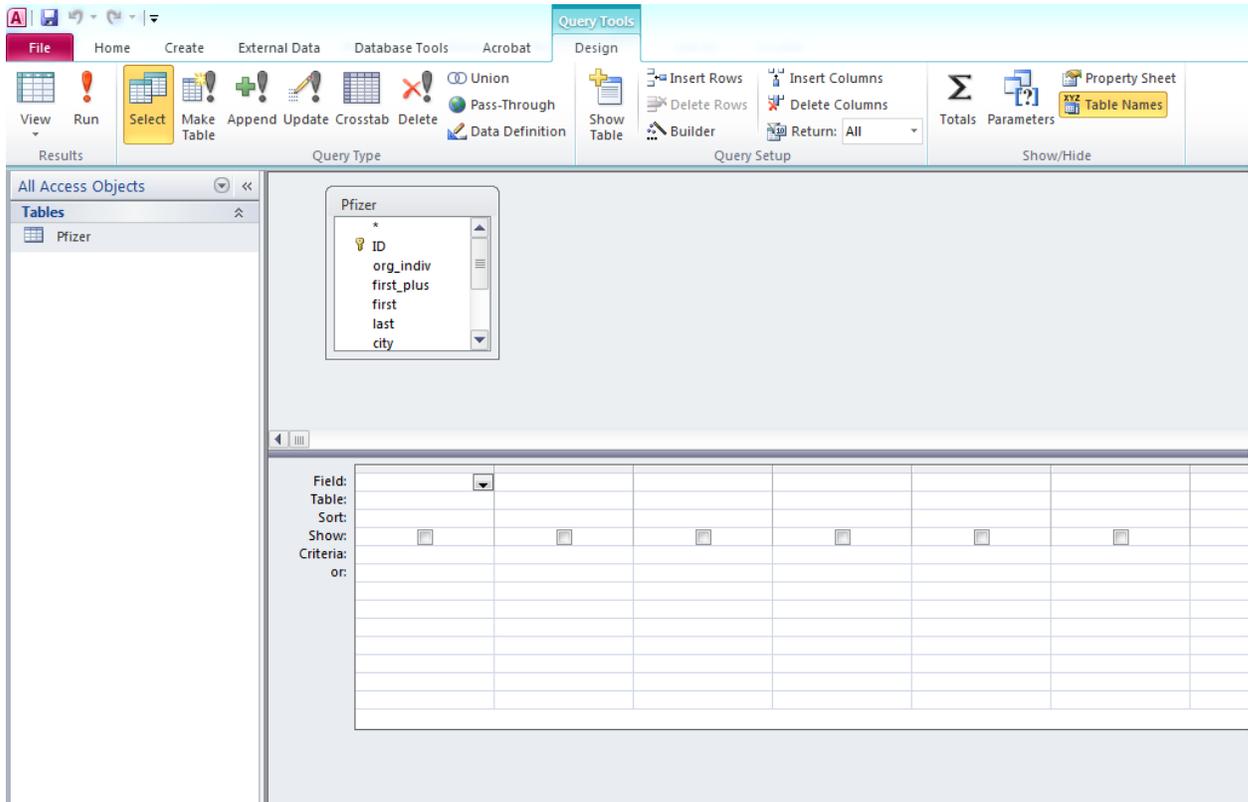
See that it looks much the same as a spreadsheet, except there is no coordinate system. Instead, the column names, called “fields” in a database, are fixed, and each row or “record” has an ID number, created by Access as a “Primary Key” when the data was imported. (We’ll do this with a new table in a few minutes.)

Notice also that the field names are simplified and have no spaces – we’ll see why later on.

## 1. Filtering queries

Let's repeat our spreadsheet task of making a list of all doctors in California who were paid \$10,000 or more to run expert-led forums.

Select **Query Design** under the **Create** tab. Then **Add** the table and the screen should look like this:



Click on **Field** in the first column of the grid, and select **first\_plus**. Carry along across the grid, until the following fields have all been selected:

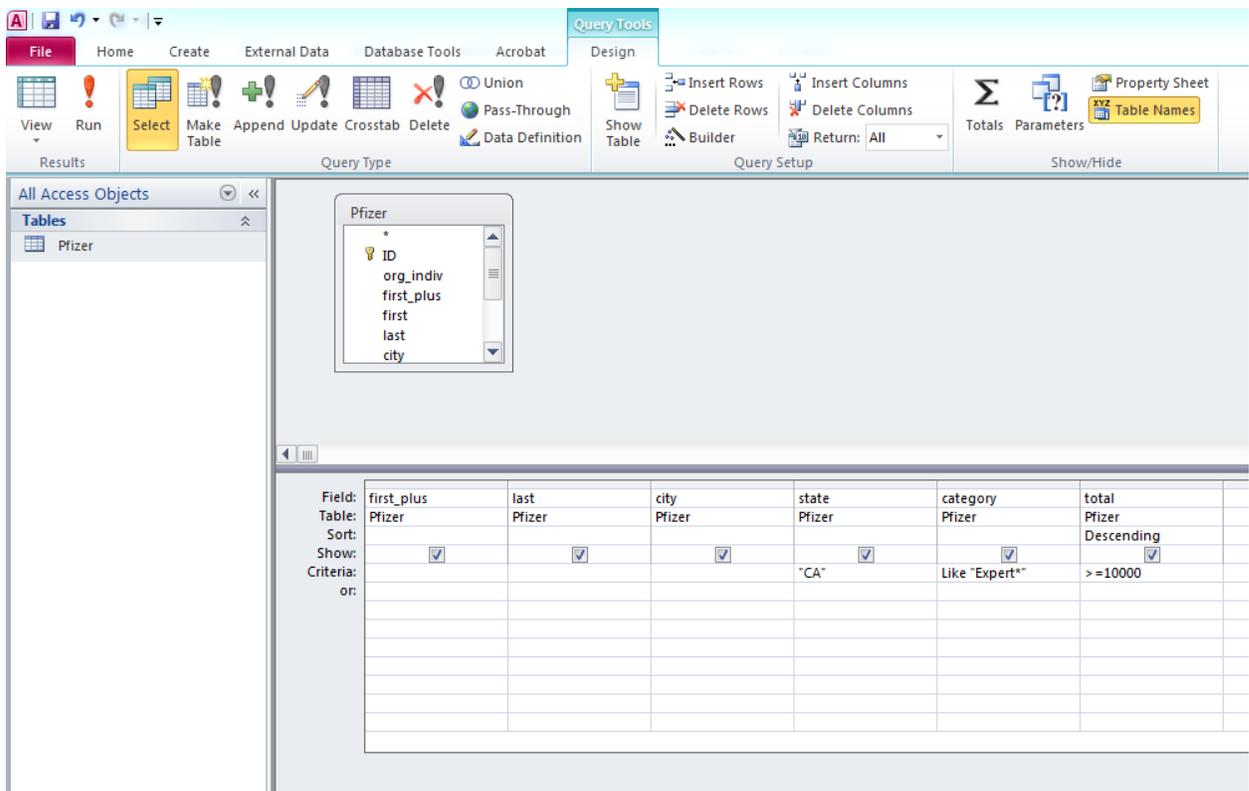
The screenshot shows the Microsoft Access Query Design View. The ribbon includes 'Query Tools' with 'Design' and 'Query Setup' tabs. The 'Pfizer' table is selected in the 'All Access Objects' pane. The field list for 'Pfizer' includes ID, org\_indiv, first\_plus, first, last, and city. The query grid below shows the following configuration:

Field:	first_plus	last	city	state	category	total
Table:	Pfizer	Pfizer	Pfizer	Pfizer	Pfizer	Pfizer
Sort:						
Show:	<input checked="" type="checkbox"/>					
Criteria:						
or:						

Now we will filter using **Criteria**: type **CA** in the column for **state** and **Like Expert\*** under **category**, and **>=10000** under **total**. Also in this column, click on **Sort** and select **Descending**.

**Tip!** Like **Expert\*** tells Access to look for entries that begin with **Expert**. In this case it's just saving us some typing, but filtering using this wildcard function can be helpful to return records that may have been entered into a database slightly differently. The wildcard **\*** can be inserted anywhere.)

The screen should now look like this:

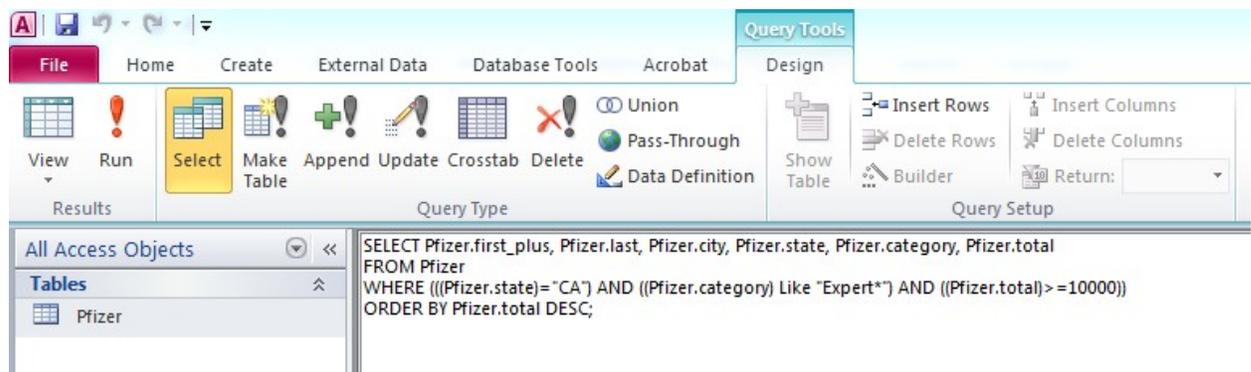


Hit **!** to run the query, and we should obtain the same list of 31 doctors we had before.

first_plus	last	city	state	category	total
GERALD MICHA	SACKS	SANTA MONIC	CA	Expert-Led For	\$146,500.00
MITCHELL	NIDES	LOS ANGELES	CA	Expert-Led For	\$70,500.00
STEVEN GARTH	POTKIN	ORANGE	CA	Expert-Led For	\$48,350.00
DAVID ALAN	GINSBERG	LOS ANGELES	CA	Expert-Led For	\$45,750.00
SAMUEL	LOUIE	SACRAMENTO	CA	Expert-Led For	\$41,250.00
GURKIPAL	SINGH	WOODSIDE	CA	Expert-Led For	\$40,000.00
IVAN STEPHEN	BAROYA	BONITA	CA	Expert-Led For	\$26,400.00
MATTHEW JAY	BUDOFF	MANHATTAN E	CA	Expert-Led For	\$24,000.00
QUANG H	NGUYEN	LA JOLLA	CA	Expert-Led For	\$22,500.00
JOHN SPEER	SCHROEDER	STANFORD	CA	Expert-Led For	\$21,500.00
DANIEL SHAHR	BANDARI	LOS ANGELES	CA	Expert-Led For	\$21,000.00
ANDREW M	BLUMENFELD	DEL MAR	CA	Expert-Led For	\$20,500.00
BRIAN RANDAL	KAYE	BERKELEY	CA	Expert-Led For	\$18,000.00
GARY WILLIAM	WILLIAMS	LA JOLLA	CA	Expert-Led For	\$18,000.00
SHAGUN	CHOPRA	SAN DIEGO	CA	Expert-Led For	\$17,250.00
FAIROOZ F	KABBINAVAR	LOS ANGELES	CA	Expert-Led For	\$17,250.00
GREGG CURTIS	FONAROW	LOS ANGELES	CA	Expert-Led For	\$15,000.00
YUNGAE KRIST	KIM	LOS ANGELES	CA	Expert-Led For	\$14,000.00
TAKKIN	LO	LOMA LINDA	CA	Expert-Led For	\$13,625.00
MICHAEL JAME	HARBOUR	PALO ALTO	CA	Expert-Led For	\$13,500.00
MARK STEVEN	WALLACE	LA JOLLA	CA	Expert-Led For	\$13,500.00
RICHARD	CASABURI	RANCHO PALO	CA	Expert-Led For	\$13,000.00
GLENN RICHA	EHRESMANN	LOS ANGELES	CA	Expert-Led For	\$12,000.00
EMILY ELIZABE	COLE	SAN DIEGO	CA	Expert-Led For	\$12,000.00
SCOTT LEE	ZELLER	ORINDA	CA	Expert-Led For	\$11,500.00
PAUL N	BARKOPOULOS	LOS ANGELES	CA	Expert-Led For	\$11,500.00
ALEX JAVIER	KOPELOWICZ	GRANADA HILL	CA	Expert-Led For	\$11,500.00
BENJAMIN JESS	ANSELL	IRVINE	CA	Expert-Led For	\$11,250.00
CLIFFORD KEITH	BECK	TORRANCE	CA	Expert-Led For	\$10,500.00
SAMUEL CRAIG	RISCH	SAN FRANCISC	CA	Expert-Led For	\$10,500.00
WILLIAM DAVII	HARDY	LOS ANGELES	CA	Expert-Led For	\$10,000.00
*					

**Ctrl-S** to save the query, giving it an appropriate name.

Click on **SQL** at the bottom-right of the screen and we can see how Access interprets the query, in **Structured Query Language**:



This can be written more simply than Access's translation from the Design View:

```
SELECT first_plus, last, city, state, category, total
FROM Pfizer
WHERE state="CA" AND category Like "Expert*" AND total>=10000
ORDER BY total DESC;
```

Now it should be obvious why databases use abbreviated field names: it makes writing SQL queries much simpler!

Learning SQL is a good idea, as with practice it is quicker than using the Design View, and is a skill that transfers to other database managers.

**TIP!** Even when working in Design View, if you are running several related queries, you don't need to laboriously fill in the grid in Design View each time. Switch to the SQL View, copy it into a new query, then switch back to the Design View and edit from there.

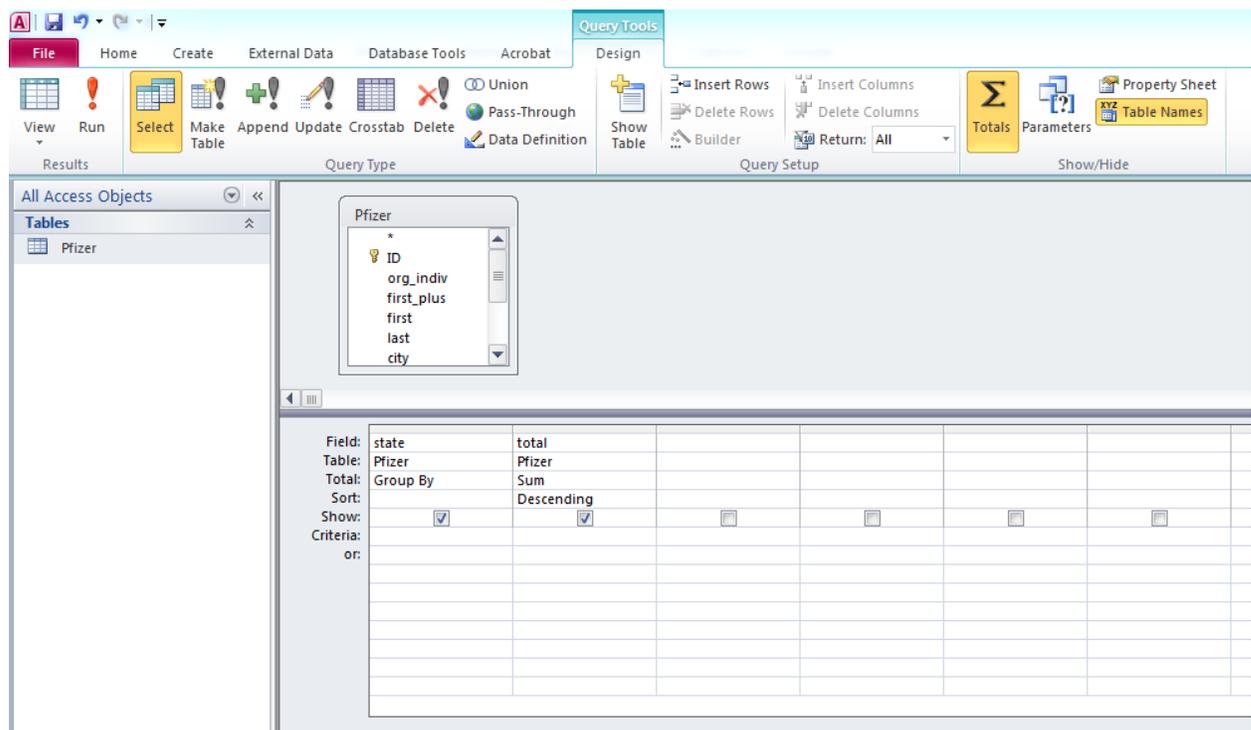
## 2. Grouping and running calculations on the data

First, let's subtotal the payments by state, as before.

Create a new query, **Add** the table and click the  $\Sigma$  button, which adds **Total** to the grid. This is where we will group records together and perform calculations on them.

Select **state** in the first column of the grid, and accept the default function of **Group By**. Then add **total** in the second column, this time change the function to **Sum**. Also in this column, click on **Sort** and select **Descending**.

The screen should now look like this:



Here is the same query in SQL (giving a new name for the calculated field):

```
SELECT state, Sum(total) AS state_total  
FROM Pfizer  
GROUP BY state  
ORDER BY Sum(total) DESC;
```

Hit **!** to run the query (notice how in Design View Access gives a default name to the calculated field):

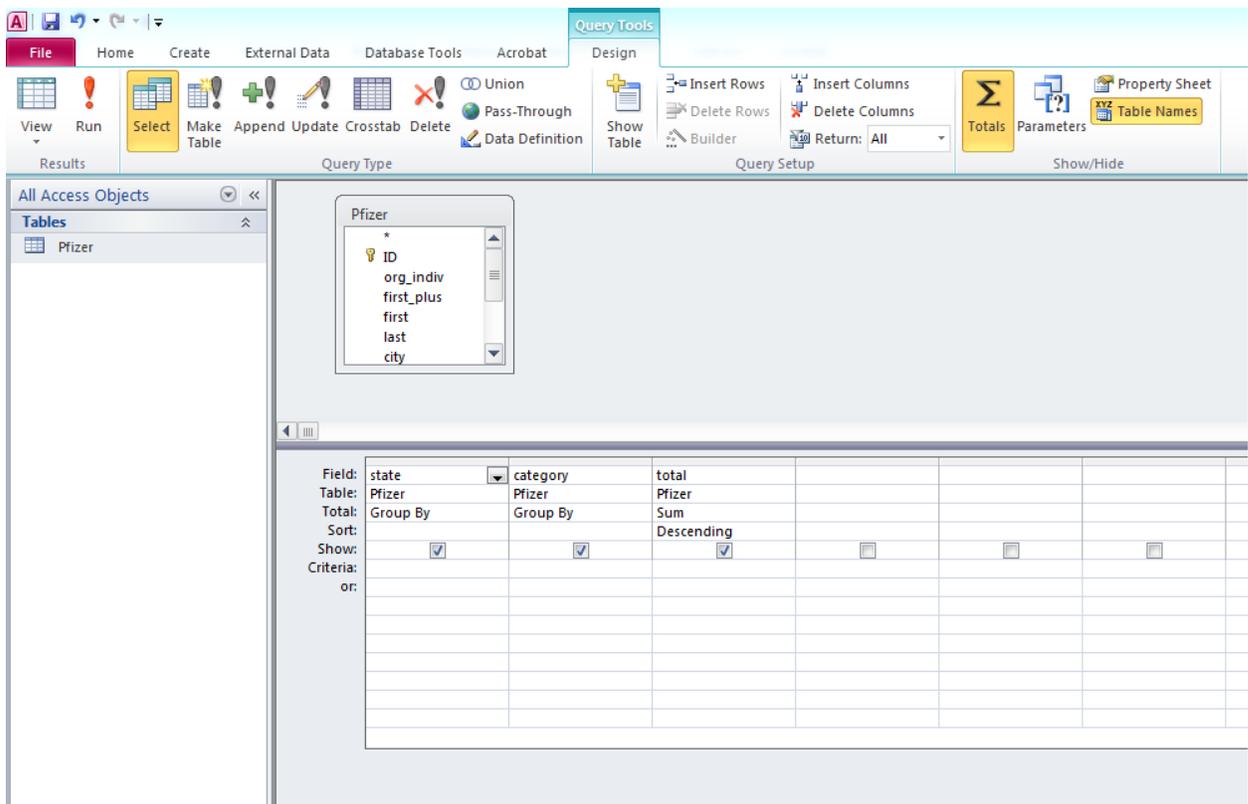
state	SumOftotal
CA	\$4,737,807.00
TX	\$2,802,196.00
FL	\$2,564,047.00
PA	\$2,484,505.00
NC	\$2,328,435.00
NY	\$2,065,042.00
MA	\$1,764,771.00
IL	\$1,256,825.00
MI	\$1,146,285.00
OH	\$1,019,450.00
MO	\$973,586.00
CO	\$915,238.00
MD	\$870,905.00
TN	\$849,225.00
AL	\$681,699.00
AZ	\$641,851.00
CT	\$632,282.00
GA	\$618,645.00
NJ	\$600,842.00
MN	\$569,300.00
WI	\$510,122.00
KY	\$436,938.00
SC	\$421,491.00
WA	\$396,066.00
UT	\$380,892.00
VA	\$367,992.00
IN	\$349,589.00
KS	\$307,205.00
OR	\$303,740.00
LA	\$261,921.00
DC	\$250,541.00
IA	\$243,706.00
RI	\$210,204.00
NE	\$200,250.00
NH	\$172,369.00
AR	\$160,932.00
PR	\$130,394.00
WV	\$128,372.00
OK	\$111,523.00
MS	\$85,276.00
NV	\$73,024.00
NM	\$63,830.00
DE	\$53,987.00
HI	\$42,617.00
WY	\$39,962.00
ID	\$37,656.00
VT	\$29,888.00
SD	\$29,503.00
ME	\$18,731.00
ND	\$16,146.00
MT	\$11,208.00
AK	\$1,750.00

Save the query as before.

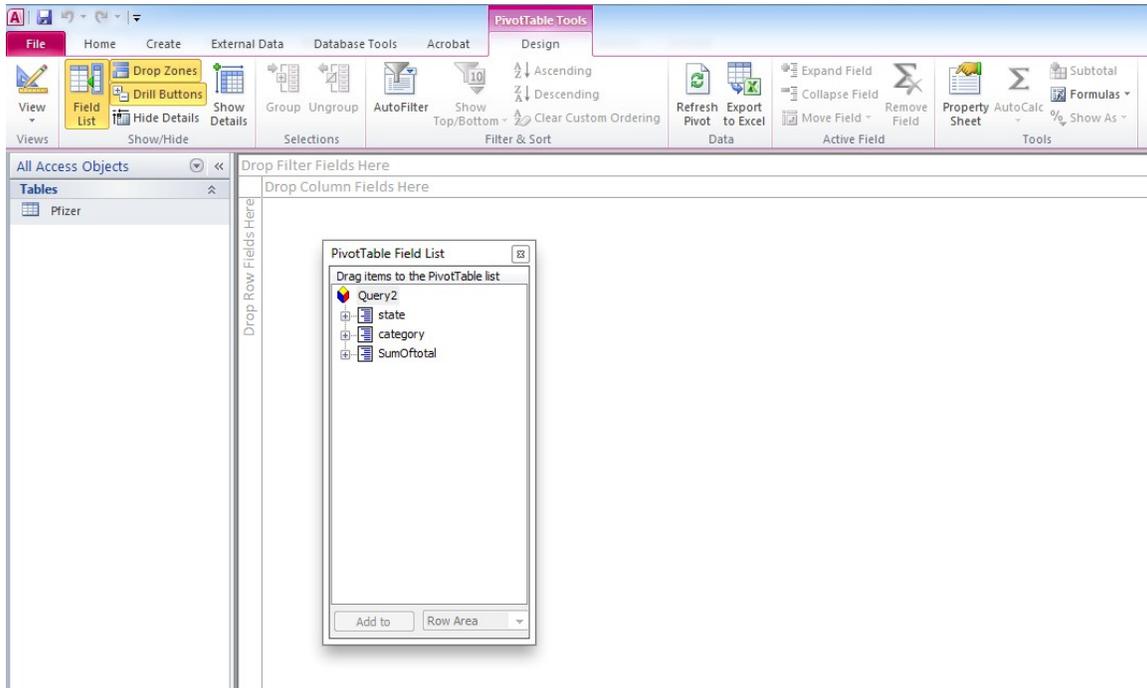
**TIP!** To export a saved query, select **Export to Excel** or **Export to Text File** under the **External Data** tab.

Now let's create a pivot chart, totaling the payments by state and category simultaneously.

Set up the query as before, but this time Grouping by **category** as well as by **state**, as follows:



Then select the **Pivot Table View** button at bottom right, and the screen should look like this:



Add **state** to the **Column Area**, **category** to the **Row Area** and **SumOfTotal** to the **Data Area**, and this should be the result:

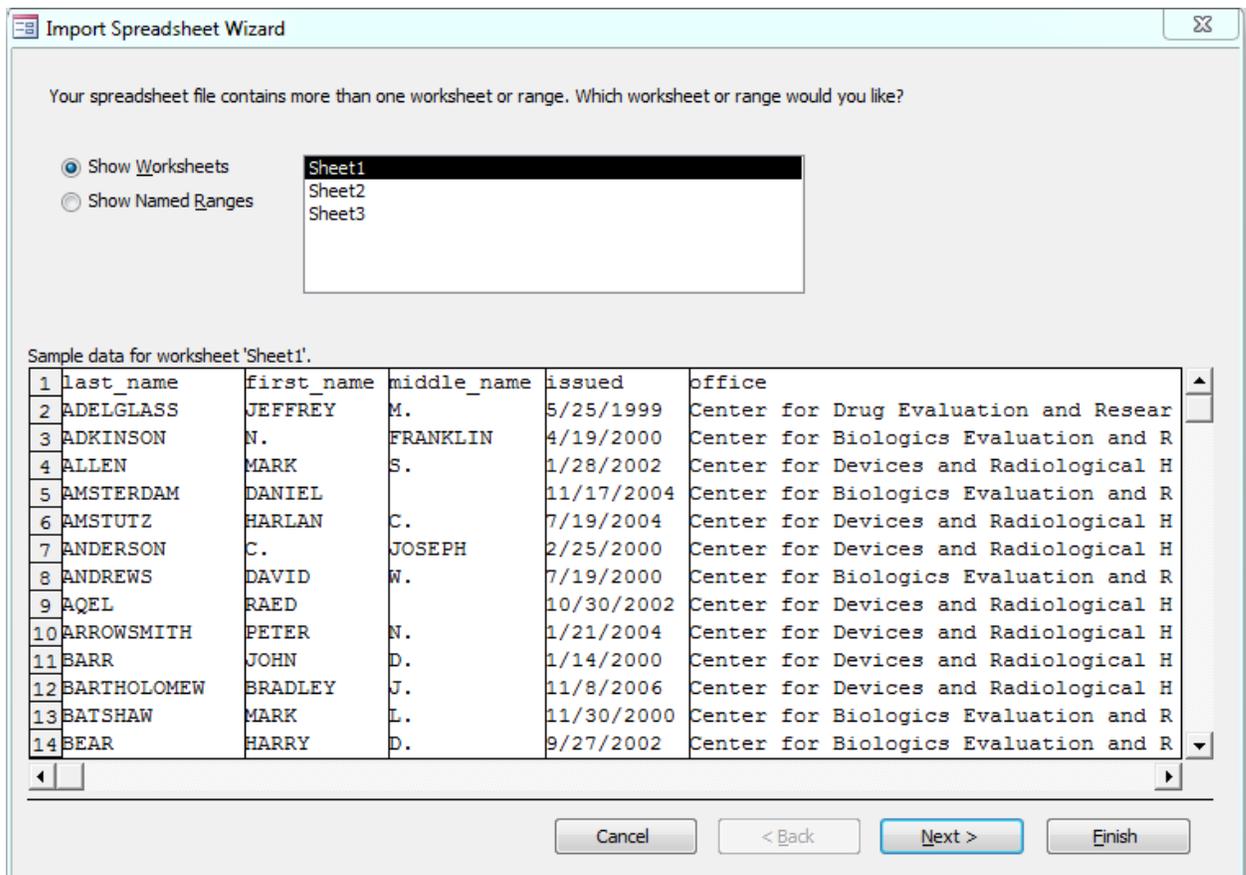
The screenshot shows the PivotTable view of the data. The PivotTable is structured with 'state' as the column field, 'category' as the row field, and 'SumOfTotal' as the data field. The data is summarized by state (AK, AL, AR, AZ, CA, CO) and category.

category	state					
	AK	AL	AR	AZ	CA	CO
(Blank)						
Business Related Travel		\$29,123.00	\$3,783.00	\$15,101.00	\$248,627.00	\$18,207.00
Educational Items		\$431.00	\$318.00	\$173.00	\$947.00	\$144.00
Expert-Led Forums	\$1,750.00	\$129,850.00	\$72,825.00	\$134,600.00	\$1,460,650.00	\$176,550.00
Investigator-Initiated Research		\$55,937.00		\$76,102.00	\$173,563.00	
Meals		\$10,217.00	\$3,700.00	\$13,409.00	\$156,716.00	\$12,433.00
Pfizer Sponsored Research initiated before July 1, 2009		\$279,485.00	\$77,670.00	\$225,615.00	\$1,125,624.00	\$596,437.00
Pfizer Sponsored Research initiated on or after July 1, 2009		\$66,096.00		\$107,979.00	\$328,149.00	
Professional Advising		\$110,560.00	\$2,636.00	\$68,872.00	\$1,243,531.00	\$111,467.00
Grand Total	\$1,750.00	\$681,699.00	\$160,932.00	\$641,851.00	\$4,737,807.00	\$915,238.00

## Matching and searching across multiple data tables

### 1. Importing a new data table

Select **Import Excel spreadsheet** under the **External data** tab. Browse for the file **FDA warning letters clinical investigators.xls**, select **Import the source data into a new table in the current database**



Click the **Next** button.

At the next dialog box, make sure to check **First row contains column headings**.

Import Spreadsheet Wizard

Microsoft Access can use your column headings as field names for your table. Does the first row specified contain column headings?

First Row Contains Column Headings

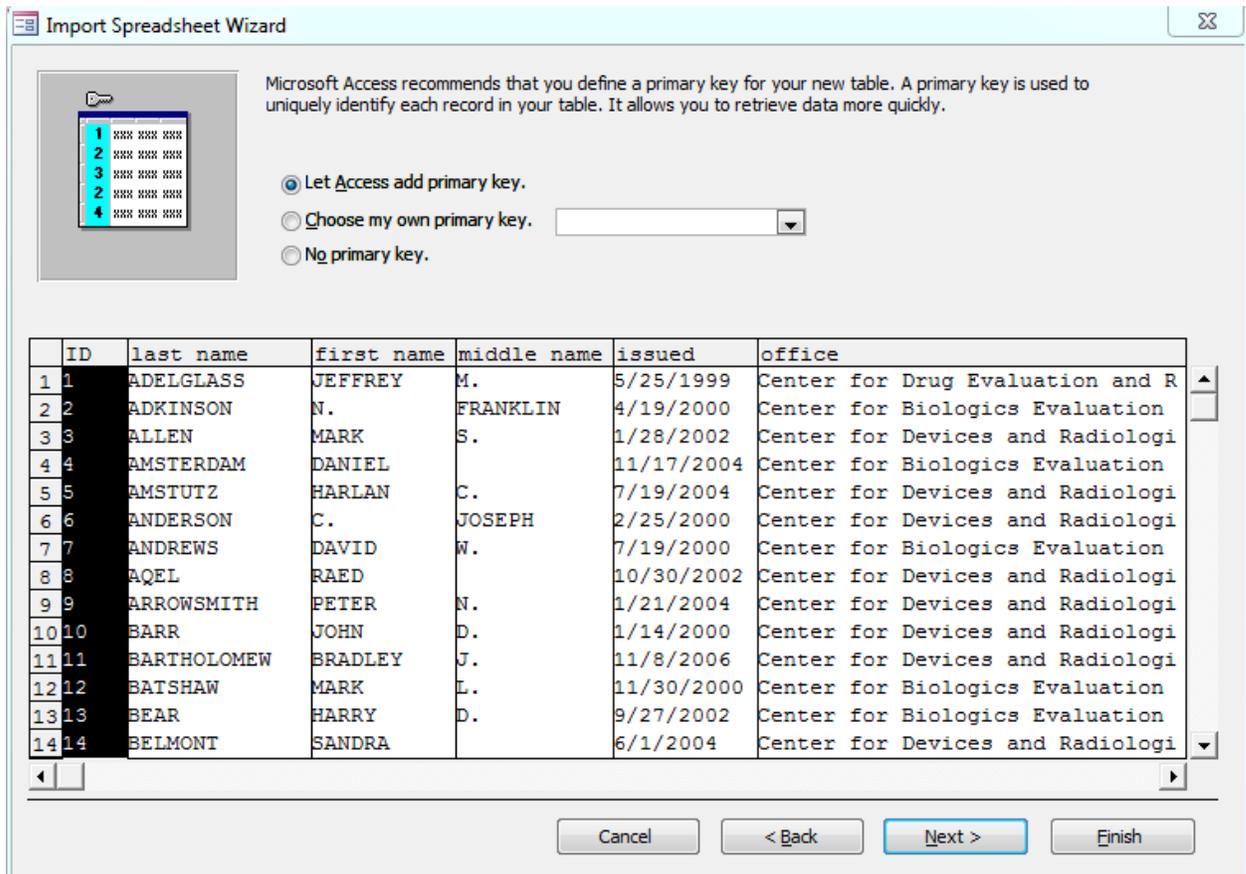
	last name	first name	middle name	issued	office
1	ADELGLASS	JEFFREY	M.	5/25/1999	Center for Drug Evaluation and Resear
2	ADKINSON	N.	FRANKLIN	4/19/2000	Center for Biologics Evaluation and R
3	ALLEN	MARK	S.	1/28/2002	Center for Devices and Radiological H
4	AMSTERDAM	DANIEL		11/17/2004	Center for Biologics Evaluation and R
5	AMSTUTZ	HARLAN	C.	7/19/2004	Center for Devices and Radiological H
6	ANDERSON	C.	JOSEPH	2/25/2000	Center for Devices and Radiological H
7	ANDREWS	DAVID	W.	7/19/2000	Center for Biologics Evaluation and R
8	AQEL	RAED		10/30/2002	Center for Devices and Radiological H
9	ARROWSMITH	PETER	N.	1/21/2004	Center for Devices and Radiological H
10	BARR	JOHN	D.	1/14/2000	Center for Devices and Radiological H
11	BARTHOLOMEW	BRADLEY	J.	11/8/2006	Center for Devices and Radiological H
12	BATSHAW	MARK	L.	11/30/2000	Center for Biologics Evaluation and R
13	BEAR	HARRY	D.	9/27/2002	Center for Biologics Evaluation and R
14	BELMONT	SANDRA		6/1/2004	Center for Devices and Radiological H

Cancel < Back Next > Finish

Click Next twice (in this case Access will recognize the correct data types, but normally you will want to check and correct if necessary).

**Tip!** If you get import errors, re-enter the data selecting the column with problems as text, then change the data type for that column once you have created the table.

At the dialog box illustrated below, select **Let Access Add primary key**, then click **Finish**, naming the table **FDA**.

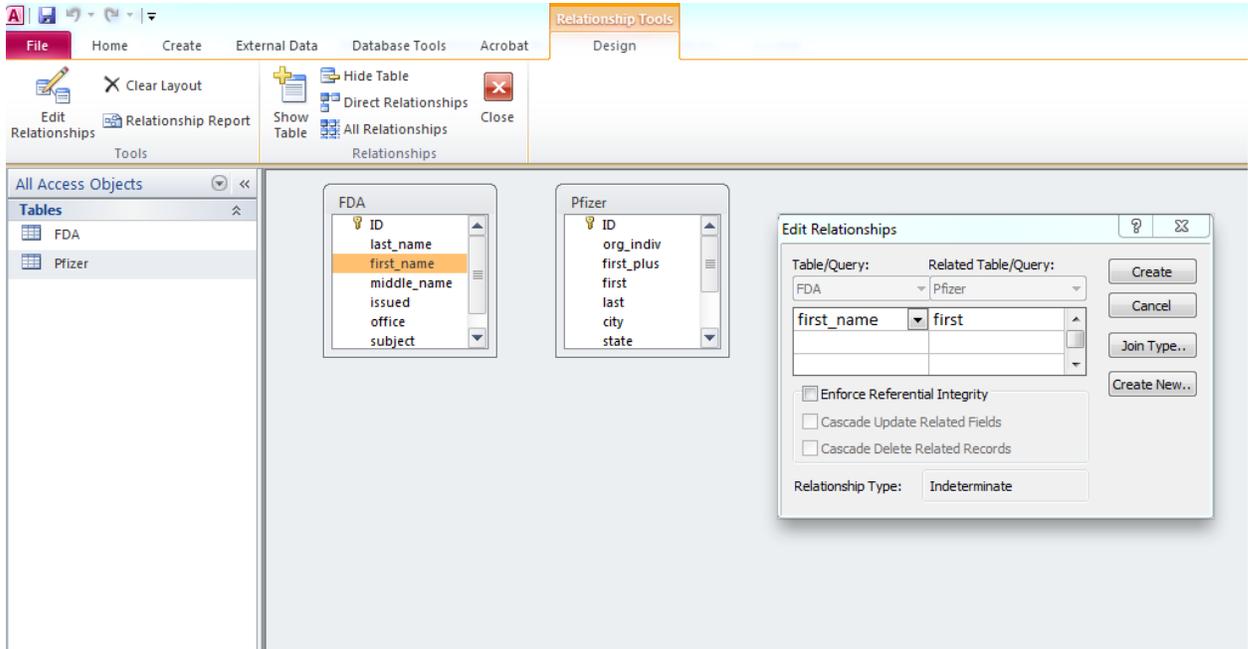


## 2. Joining data tables

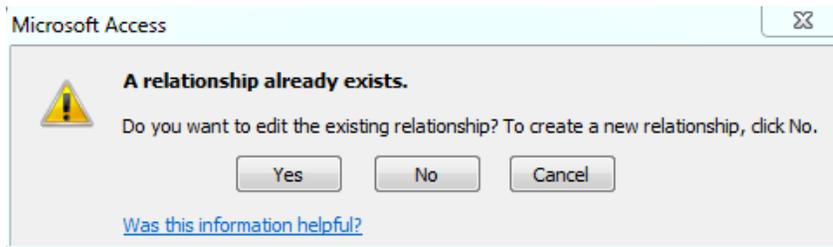
Now we're going to create a query across the two data tables, having linked them together in the right way, so we select doctors paid by Pfizer to run Expert-led forums who had also received a warning letter from the FDA for problems in their conduct of clinical research.

To link the tables, select **Relationships** under the **Database Tools** tab, select and **Add** both tables, then close the dialog box.

Now click on **first\_name** in the FDA warning letters table and drag it across to **first** in the Pfizer payments table. When the dialog box pops up, click on **Create**:

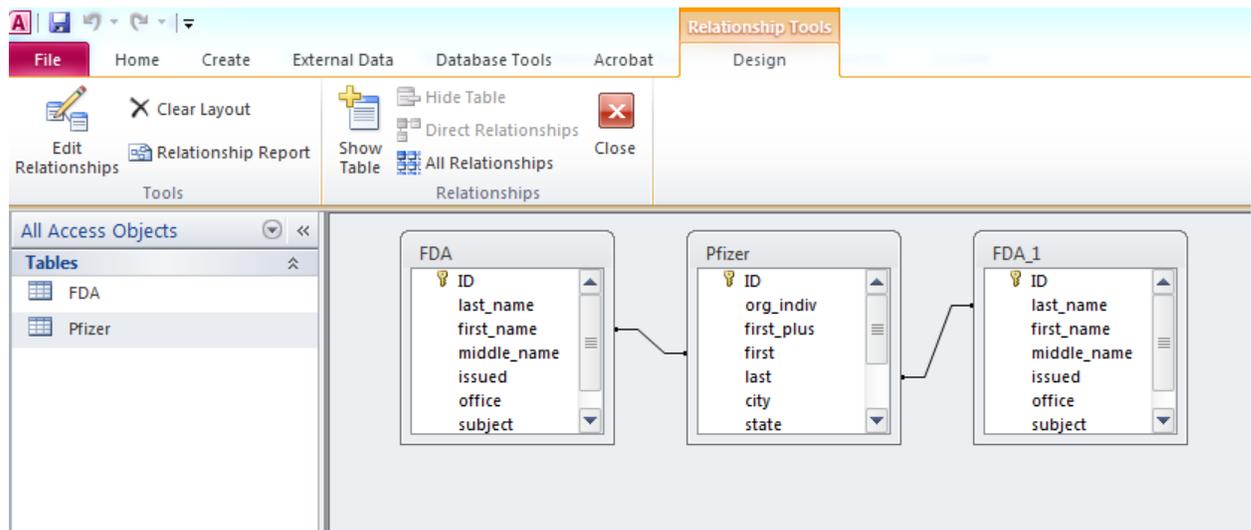


Repeat the process for the **last\_name** and **last** fields, clicking **No** when you see this dialog box:



(This has created **INNER JOINS** between the two tables, so that queries run across them will only return data if the first and last name fields match. It is also possible to link tables in other ways, using the **Join Type** button in the dialog box above. For example, queries run on **LEFT** or **RIGHT JOINS** would return all the requested data from the one table, and just those with matching first and last names from the other. )

The screen should now look like this:

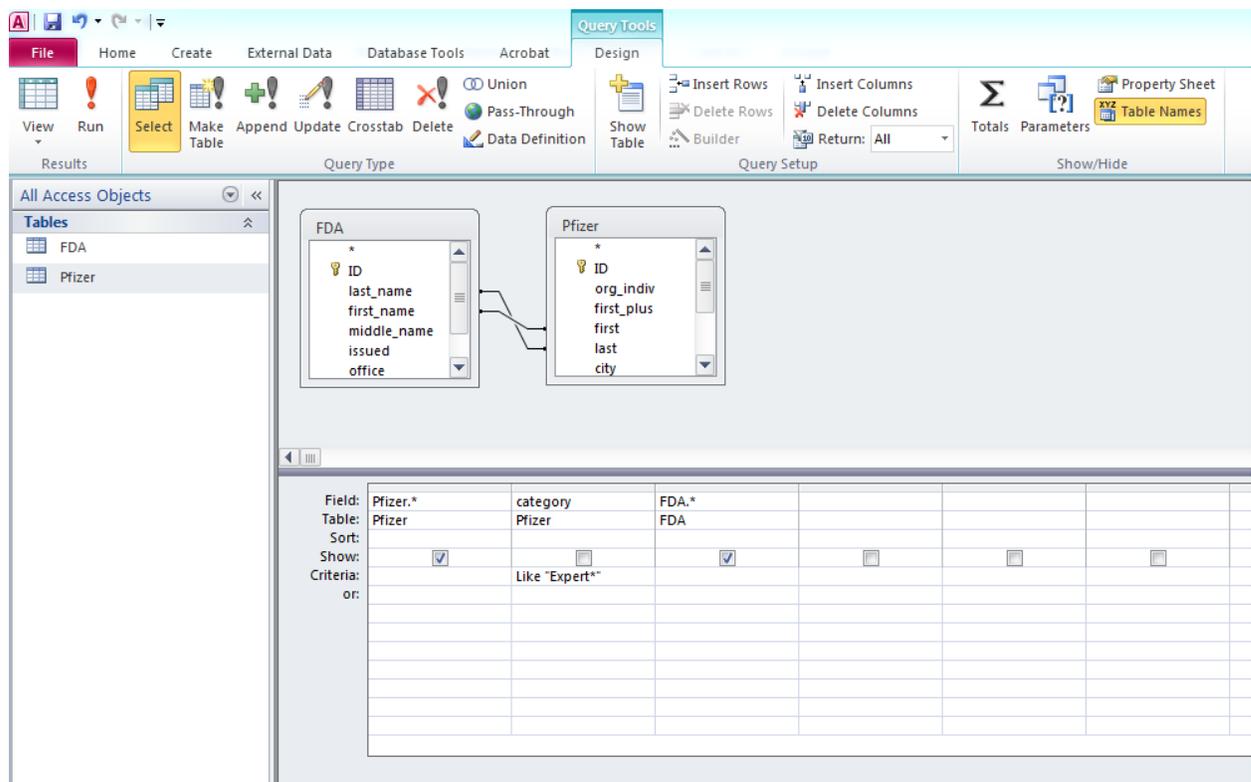


**Ctrl-S** to save the new relationship and close.



Fill in the grid as follows. Selecting \* returns all relevant records from a table. Adding **category** from the Pfizer table allows us to filter to just see those doctors paid to run expert forums, using Criteria: **Like Expert\***. Uncheck **Show** so that **category** doesn't appear twice in the results.

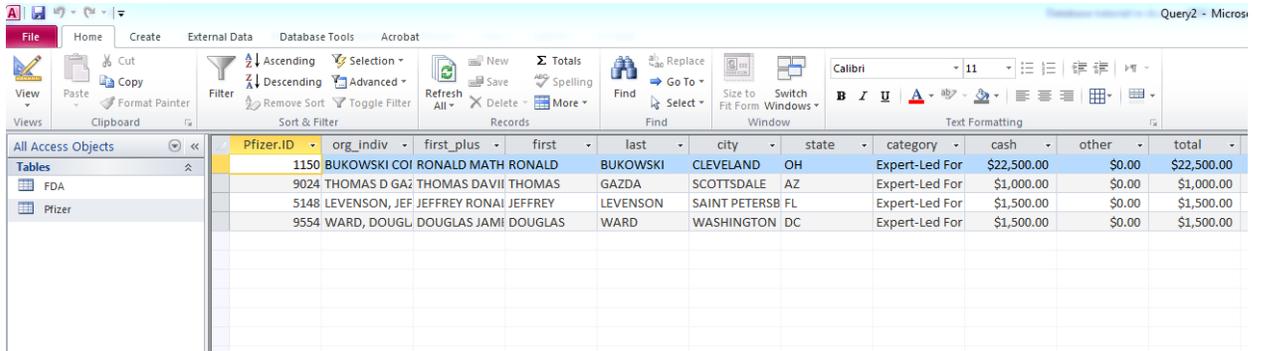
The screen should now look something like this:



Here is the SQL for this query:

```
SELECT Pfizer.*, FDA.*
FROM FDA INNER JOIN Pfizer ON FDA.last_name = Pfizer.last AND
FDA.first_name = Pfizer.first
WHERE Pfizer.category Like "Expert*";
```

Hit ! to run the query, and we should have four doctors:



The screenshot shows the Microsoft Access interface with a query result table. The table has the following columns: Pfizer.ID, org\_indiv, first\_plus, first, last, city, state, category, cash, other, and total. The data is as follows:

Pfizer.ID	org_indiv	first_plus	first	last	city	state	category	cash	other	total
1150	BUKOWSKI COI	RONALD MATH	RONALD	BUKOWSKI	CLEVELAND	OH	Expert-Led For	\$22,500.00	\$0.00	\$22,500.00
9024	THOMAS D GA2	THOMAS DAVII	THOMAS	GAZDA	SCOTTSDALE	AZ	Expert-Led For	\$1,000.00	\$0.00	\$1,000.00
5148	LEVENSON, JEF	JEFFREY RONAI	JEFFREY	LEVENSON	SAINT PETERSB	FL	Expert-Led For	\$1,500.00	\$0.00	\$1,500.00
9554	WARD, DOUGL	DOUGLAS JAMI	DOUGLAS	WARD	WASHINGTON	DC	Expert-Led For	\$1,500.00	\$0.00	\$1,500.00