Transcript

Epi Info™ 7 Instructional Video

Overview (00:06)

Hello, and welcome to this demonstration of Epi Info 7. My name is Erik Knudsen, and I'm an Information Technology Specialist with the Epidemiology and Analysis Program Office, part of the Centers for Disease Control and Prevention.

Agenda (00:27)

Before I begin, I would just like to go over a very quick agenda. I'm first going to talk a little bit about what Epi Info is, for those that might not be familiar, and then I will discuss some core design principles that have gone into the new version, and then finally I'm going to give a demonstration of some of the features and capabilities found in Epi Info 7.

Epi Info Overview (00:45)

So What is Epi Info? (00:50)

Epi Info is a suite of free data management, analysis, and visualization tools designed specifically for the public health community. It is used extensively, not just at CDC, but also domestically at the state and local level, as well as internationally. Epi Info allows you to:

Rapid Electronic Form Creation and Data Entry (01:11)

Rapidly create electronic data entry forms; these forms can have intelligence built into them. For example, you can automatically calculate a patient's age based on the survey data date of birth; you can also move the cursor past certain fields; you can even hide and unhide fields based on various conditions. It also allows you to enter data into those forms.

Statistical Analysis (01:42)

You can conduct various types of statistical analysis, including frequencies, 2x2 (two by two) tables, analysis of variance, conditional and unconditional logistical regression, Kaplan Meier survival analysis, complex samples, and many others.

Mapping and Visualization (02:00)

Epi Info also allows you to create several types of maps, and also has graphing and charting capabilities.

Report Design and Generation (02:10)

And finally, Epi Info allows you to design and generate reports.

Current Usage (02:15)

We have found that Epi Info currently averages 227 (two hundred and twenty seven) downloads per day. The vast majority of these downloads come from the United States, and we have registered at least one download from every one of the 50 state health departments. Now, when users download Epi Info from our website, they are given the option of telling us how many people they are downloading it for, and 7 (seven) percent of those users are telling us they are downloading Epi Info for groups of 10 (ten) or more users.

Epi Info 7 Core Principles (02:57)

We have developed Epi Info 7 to adhere to a number of design principles.

Free (03:08)

First among those is that we want Epi Info to remain free. And when we say free, we don't just mean free to download, but we also mean free as in you won't need to buy any other software in order to get Epi Info to work. You'll be able to download Epi Info from our website, and begin running it without having to purchase any additional software.

Easy to Use (03:32)

We also want Epi Info 7 to be easy to use. We want our users to be able to very quickly and easily figure out how to do certain tasks, and from the previous version of Epi Info we want to make sure that the number of steps needed to do certain tasks is reduced.

Flexible (03:50)

We also want Epi Info 7 to be very flexible, and when we say flexible, we really mean two things. Number one, we want Epi Info to be lightweight and agile, especially when responding to emergencies. So that for example, if you are deployed to a disaster area, where there might not be any internet or infrastructure, we want Epi Info 7 to be usable in that type of scenario, so that you can simply take it with you and begin working on whatever computers you happen to find on site.

(04:33)

Number two, at the same time, we want Epi Info to be very robust, and what we mean by that is that it can work with very large sets of data. So we've designed Epi Info 7 to work not only with the traditional Microsoft Access database format, but it can now work with Microsoft Sequel Server databases. What this allows us to do is to work with gigabytes of data, thousands of fields, and even millions or tens of millions of records.

Standards Based (05:10)

We also want Epi Info 7 to be standards based. To that end we have incorporated the PHIN VADS vocabulary system into the Epi Info 7 form designer. What that lets our users do is create their questions as standard public health responses, so that when the data is transferred, you know that the data has been stored in a standard way.

No "IT Guys" Needed (05:42)

Last but not least, we've tried to eliminate the amount of IT support needed. One of the things we've done with Epi Info 7 is removed the requirement that it be installed to the computer. In fact, all you have to do to get Epi Info to work is download it from the website and then run it. You won't need an administrator, you won't need your IT support personnel to set it up for you, or install it for you. Just download it and get right to work. In fact, you can download Epi Info 7 directly to a thumb drive and take that thumb drive with you and wherever you happen to be, you can plug that thumb drive into a Windows computer and simply run Epi Info 7 right from the thumb drive.

(06:37)

That concludes the Power Point demonstration.

<u>**Epi Info 7 Demonstration (06:38)**</u>

Let's get right to the demo. I'm going to launch Epi Info 7 right from the desk top.

(07:05)

[Epi Info splash screen appears with top level menu choices: Create Forms, Enter Data, Analyze Data: Classic and Analyze Data: Visual Dashboard, Create Maps, Epi Info Website, and Exit].

What we're looking at now is the Epi Info 7 main menu. From this screen I can create electronic forms, enter data into those forms, do various types of analysis, and I can work with maps.

(07:12)

Let's start out with one of the forms we've already created, and entered some fake data into.

(07:19)

[Enter Data screen appears, File menu is selected, then Recent Forms selected to show a list of three forms; presenter choose eColi\FoodHistory form]

This is an E. Coli food history questionnaire; as you can see, we are collecting some pretty standard information: Case ID, the date the patient was interviewed, we have first name, last name, the patients' gender, date of birth.

(07:40)

If I scroll down to the bottom half of the page, you can see that there are some additional questions here on symptoms and illness: whether the patient was hospitalized, whether they were treated with antibiotics, and whether they died.

(07:57)

There's a second page to this questionnaire which I can navigate to by clicking on page two [Page 2] in the list of pages [in the menu at the upper left].

(08:06)

On page two, we now have a group box of various foods that the patient might have eaten.

(08:17)

[Instructor scrolls down to bottom of this page] Here we have some lab tests and lab result questions as well.

 [Instructor clicks again on Page one from the Page List in the upper left corner menu]. I can navigate through the patient records by clicking the Next Record [instructor clicks on the icon of vertical line and right-pointing arrow in the top menu bar] and Previous Record buttons [instructor clicks on the icon of the left-pointing arrow in the top menu bar] and it's also showing me that I have three hundred and fifty nine total records [instructor's mouse hovers over the Record Count in the top menu bar].

One thing that I mentioned at the very beginning is that Epi Info 7 questionnaires have intelligence built into them. So for instance, I skip past all questions related to pregnancy if I filled out Male for the gender. Now in this particular form, we have programmed it so that the cursor, as soon as it moves to the Date of Birth [DOB] field, it will automatically calculate the patients' age based on the Date of Interview and the Date of Birth, and then automatically fill in that value to the Age field. As soon as I put the cursor into the Date of Birth field, then press Tab, you'll notice that the Age field was filled in with twenty nine [29]. If I were to go back and change the Date of Birth field to January twelfth, two thousand and one [01-12-2001], and then press Tab, we see that the Age is recalculated as ten [10].

I can create a Line Listing right from within the Enter module [instructors' mouse moves to Line Listing in the top menu bar]. We're going to do that by clicking Line Listing at the top of the screen. It's going to give me several options to display this line list [choices shown in pulldown menu are: Interactive, Printable (HTML), and MS Excel]. I can create an Interactive line list, which would allow me to create various data filters, I can generate a Printable line list, which will show up in whatever web browser is the default on this computer, or I can export that line list directly to Microsoft Excel. We're going to go ahead and do the printable line list, and in a couple of seconds I can see all of my data in Internet Explorer and I can scroll down and scroll across [scrolls down, across, and back up to the top of the list of all patient records].

Now, one of the new features that we've added to Epi Info 7 is called the Dashboard, and the Dashboard is really just a quick way for users to see what kinds of data they are collecting, as the data is being collected. So I'm going to click this Dashboard button at the very top of the screen, and the Dashboard window appears. It's showing me that I have three hundred and fifty nine records, and it's also telling me to Right-click on the canvas to add content. So if I right-click, it's now going to give me the option of adding several different gadgets [shows list of Add Analysis Gadget, Add StatCalc Calculator, Add NutStat Growth Chart, with further pulldown choices for each of those].

Of the several different Analysis Gadgets I can add a two by two table, a combined frequency gadget, we can also add gadgets for linear and logistic regression, and various kinds of charts. [Instructor displays further pulldown menu choices for the Add Analysis Gadget are Frequency, EARS Aberration, Combined Frequency, 2 x 2 Table, M x N Table, Means, Linear Regression, Logistic Regression, and Chart]

I can add various types of StatCalc calculators. [Instructor displays further pulldown menu choices for the Add StatCalc Calculator: Tables (2 x 2, 2 x n), Sample size and power, and Chi square for trend].

I can also add various NutStat growth charts. [Instructor displays further pulldown menu choices for the Add NutStat Growth Chart: CDC 2000 Growth Reference, WHO Child Growth Standards, WHO Reference 2007, CDC/WHO 1978].

I'm going to go ahead and enter a Frequency gadget. This gadget is going to ask me what field I want to run the frequency on. Suppose I want to see the frequency of ill patients and not ill patients in this set of data. I can simply select the Ill field – one of the choices from the questionnaire- and as soon as I do that, it's going to show me the results – there are two hundred and seventy six ill patients, and eighty three that are not ill.

If I wanted to show the breakdown of males and females instead, I could select Sex as the field to run frequency on. It is now showing me that I have one hundred and eighty six females, and one hundred and seventy three males. I can stratify this frequency by clicking, by opening the Advanced Options panel, so we'll do a frequency on Ill, and then we'll stratify it by Sex. What we're showing now is that we have one hundred and forty seven ill females and one hundred and twenty nine ill males.

I can also create some descriptive statistics for [instructor brings up Add Analysis Gadget menu and shows the sub-menu choices] various numeric fields that I have in my questionnaire. So let's say I want to know the average age of patients in this set of data. Let's select Age as the field I want to run the statistics on, and in a couple of seconds it's going to show me the Average Age, or the Mean Age, as thirty six point five (36.5). And just like we did for the Frequency, I can also open the Advanced Options panel, and in this case, I'm going to cross-tabulate by Sex. It is now showing me that the Average Age of females is thirty five point seven five five (35.755) and the Average Age of males is thirty seven point three six six (37.366). Because we've done a cross-tabulation, we're getting some

additional statistics, in this case an analysis of variance, Bartlet's Test, and Kruskal-Wallis statistics, and various others.

I can also create a two by two [2x2] Table, and since this is an E. Coli Food History questionnaire, I'm probably going to want to select one of the foods eaten, in the Exposure Field – so let's select Strawberries, whether or not the patient ate strawberries, as the exposure, [instructor selects Strawberries from the Exposure Field pulldown menu] and whether or not the patient became ill would be the outcome [instructor selects Ill from the Outcome Field pulldown menu]. In a couple of seconds, it's going to show me some pretty standard looking two by two [2x2] results, in four colored squares, I have an odds ratio, risk ratio, P-values and various other statistics.

Let's go back to the form real quick. On Page two, you may remember that we had a number of checkboxes representing each one of the various foods the patient might have eaten. These foods are all contained in this Foods Eaten group box. If we were to go through, and do one two by two [2x2] table on each one of these checkboxes, it might take us quite some time. So what we've done with Epi Info Seven is to run the statistics by every one of the fields in a given group box, and in this case we've called the group Foods Eaten, so in the Dashboard, I'm going to select the Foods Eaten group as the exposure [instructor switches to Dashboard view and chooses Foods Eaten from the Exposure Field in Crosstab Properties] and it's now showing me a relative risk chart where every row in this table represents one of the foods that was in that Foods Eaten group. If you look very closely, you'll notice that as I highlight a row in the table, the two by two [2x2] results update to reflect that particular food. So for example, if I click on Beefjerky, it now shows Beefjerky by Ill with an odds ratio of zero point nine six [0.96]. When I click on Grapes, the statistics are now updated to show Grapes by Ill with an odds ratio of zero point five [0.5]. This chart is by default sorted by risk ratio, so you can see that Beansprouts has the highest risk ratio of three point one six six [3.166]. I can sort this by exposure, alphabetically by clicking on the Exposure heading.

I can also create several types of charts [instructor selects the Add Analysis Gadget and chooses Charts from the submenu]. So let's do an Epi Curve, based on the OnsetDate. [Epi curve bar chart appears]. So we're now looking at an epi curve, and this probably won't look like a real epi curve, since the data we're using for this demo has been randomly created. Now let me go ahead and do a Frequency on Age [instructor selects Age from the Frequency Properties pulldown menu]. Now you'll notice that when I do the Frequency on Age, I'm seeing one row in my frequency table for every single value of Age that happens to be in this set of data, which is probably not what I want to see. I would rather see this broken down into some sort of age groups or age categories, and thankfully, Epi Info Seven [7] allows us to do that using the Defined Variables gadget, which is this gadget on the left-hand side of the screen. As I move the mouse over it, it sort of pops out. I'm going to create a New Variable, and I'm going to create it with the Recoded Value, since we want to recode the variables into age, so that it shows some sort of age groups. So for the Source field, I'm going to select Age, and I'm going to call it AgeGroup. Now I could manually type in all the From and the To

values to create the various age groups, but I'm going to let Epi Info do that for me by clicking on the Full Ranges button. I'm going to select my Start value as zero [0] and my End value as sixty five [65], as I want the age range to go up to sixty five years, and I want it to increment by ten [10]. I'm going to click Okay, and those values are automatically filled in to my Recode table for me. When I click Okay again, the recode is applied.

Now when I run a frequency, I see a new variable called AgeGroup and if I simply select that, it's going to show me the breakdown of ages by the categories that I just created. I can also show those age groups graphically by creating a pie chart [instructor selects Add Analysis Gadget, displays the submenu, and selects Chart to display a color pie chart with a key showing the age ranges color coded to match] and I can move the mouse over pieces of the pie to display the actual count as well as the percentage. I can create custom titles [shows Chart Configuration popup and fills in Chart Title and Legend Title - leaves X-Axis Label and Y-Axis Label blank] and apply those to the chart.

Now this is all very nice, but what if we want to run these statistics on only a certain subset of the data? For example, what if I only wanted to generate these results for male records? We can do that using the Data Filter Gadget, which is this gadget on the right-hand side of the screen. I just have to hover the mouse over it and it pops out. It's going to ask me what field I want to create a filter for, and so, since I only want to work with male records, I'm going to select Sex [for the Field Name], then Is Equal To [for the Operator] and then sex is equal to male [M-Male for the Value], then I'm going to click the Add Filter button. As soon as I do that, all of the gadgets are automatically updated and refreshed to reflect the fact that we are now only working with one hundred and seventy three [173] male records in the dataset. I can have multiple filters active at once, so if I also wanted to filter by Age, I could say the value of Age is between fifteen and sixty five [instructor selects the Data Filter Gadget, then sets Field Name to Age, Operator to Is Between, and Value to 15 and 65]. And now when I click Add Filter, it's going to ask me how I want to join this new condition with the condition I already have. I'm going to select Add Condition with And. Just like last time, all the gadgets are going to automatically update themselves to reflect that now we are only working with one hundred and fifty three records. I want to point out that the filter criteria we've created reads like an English sentence. The value of Sex is equal to M-Male and The value of Age is between 15 and 65. We've tried to make the data filtering capabilities very easy to use and tried to make the filters themselves very easy for users to figure out.

Now this is all very nice, but what if, now that I've generated all this output, I want to send it to a column. I can do that by using the Save as HTML button at the very top of the screen. When I click it, it's going to ask me where I want to save this output. I'm going to send this to my desktop, and call it DemoOutput. Now when I navigate to the desktop [instructor minimizes Epi Info and goes to desktop, and clicks on Internet Explorer file icon] and click on the HTML file, it's going to open in Internet Explorer, and I can look at all the results that I just generated.

Now I am going back to our data entry screen [instructor re-opens Epi Info and navigates to Page 1 of the Food History questionnaire] you might have noticed that we have this Get Coordinates button.

One of the various types of intelligence that can be built into Epi Info forms is the ability to geo-code an address into latitude and longitude. That service is provided by Microsoft, and therefore requires an Internet connection. It's also configurable so if you'd rather use a different service for the geo-coding, you can do so. I also want to point out that geo-coding is entirely optional, and is not required to build forms. If you don't want it, you don't have to use it.

Now I already have three hundred and fifty nine records with latitude and longitude filled in, so what if I wanted to display these cases on a map, as a case cluster? I can do that using the Mapping function. I can click this Map button at the top of the screen [map of the earth appears with continents labeled]. I'm going to add this data layer and because my data is all in latitude and longitude I'm going to select Case Cluster [instructor clicks on Add Data Layer button at top of screen and selects the Case Cluster from the pulldown menu]. It's going to ask me if I want to use external data [pop up window appears saying Use external data?]

Now this would be really nice if I had my data in an Excel spreadsheet or a SQL Server database, but because all of my data is in Epi Info Seven, I'm just going to select the No button, and it's going to ask me for the Latitude and Longitude Fields. Since I called my Latitude and Longitude fields Latitude and Longitude, I'm simply going to select those options from the list of fields. [screen changes to show a closer view of a particular area with case clusters shown as red circles with numbers on them shown in position – instructor selects the Street view radio button in the upper right corner of the screen to show a street view with case clusters shown as red circles with numbers in them].

I can use the mouse to move the map around, and can use the mouse wheel to zoom out to get a more high level view. If I zoom all the way out, I can see that I'm looking at the three hundred and fifty nine cases, records of the selected data [shows one red circle with the number 359 in the center on the map]. I can zoom in again to get a more detailed view.

One of the things you might have noticed is that we have some single dots [without numbers in them] and we have some dots with numbers inside, representing the number of cases in that cluster. If I move the mouse over one of the case clusters, in this case it has seven, you'll notice that it flares out [the circle with seven in it changes to show seven individual circles connected with lines to the large circle] where each one of these seven flares represents one of the seven cases in that cluster.

The Maps module also has a bit of intelligence built into it, where each one of the cases displayed is linked to one of the records in the inner module. So for example, if I'm looking at this map, and I see this case down here, on the southeast border of this town, and I want to maybe investigate this case a little further, I can double click on it, and that case is loaded into the Enter module for me. [The individual case page of Food History questionnaire is displayed]. Here we can see that this is case number fifty three [53], Franziska Fischer, if you wanted to you could take a look at some of her symptoms [instructor clicks on Page 2 of the questionnaire] and perhaps some foods that she might have eaten. Again, I do want to point out that this data is entirely fictional and was created for demonstration purposes only.

Now one of the things I showed in the Dashboard was the breakdown of Ill and Not Ill patients. You might recall that we only had two hundred and seventy some Ill cases in the data set. Now what if we only wanted to show those two hundred and seventy three Ill cases on our map? We can actually use the same data filtering components in the Mapping module that we used in the Dashboard. I can do that by moving the mouse over the Map Layers Gadget at the bottom of the screen and clicking the Data Filters button. I can now select Ill is equal to Yes as my filter, and when I click on Add Filter, the map is updated to reflect that we are only showing the two hundred and seventy three ill cases.

I can also create stratified maps. Let's say I want to see male records as red dots, and female records as blue dots. Well, I already have one layer with red dots, so I can set this filter's layer to only show male cases, then I'm going to add another case cluster layer [instructor selects Add Data Layer, Case Cluster from the menu at the top of the screen]; once again I'm going to say no to Use external data? And I'm going to change this layer's point color to blue [instructor mouses over color block on the right, displays a color menu, selects blue] and I'm going to set its filter to Sex is equal to Female. When I'm done picking the Latitude and Longitude fields, it's now showing me a stratified case cluster map, where the blue dots represent female cases and the red dots represent male cases.

I can also show this data temporally, in other words, I can show the progression of cases over time. I'm going to do that by clicking this Create Time Lapse button at the top of the screen. It's going to ask me which Time Variable I want to use for the time lapse, and since we collected onset date, I'm going to select OnsetDate as my Time Variable, then I'm going to click Okay. I now have this Time Gadget at the bottom of the screen. If I click on the Play button, it's going to show me the progression of cases over time, based on the date of onset. The graph that is being generated would normally look like an epi curve, but once again, this is all fake data, so the epi curve probably won't look like a traditional epi curve. I can pause, I can move the slider to any particular point in time that I want, I can increment by one day at a time, or decrement by one day at a time, or I can simply remove the indicator. [instructor closes indicator, closes the map view, and returns to Page 1 of the Food History questionnaire case record].

One last thing to point out in the Enter Module is this Linked Records area at the bottom left corner of the screen. What this allows us to do is conduct contact tracing. In other words, I can link one record to another, and say, well, this record has been exposed from this record, or this record has been exposed from this other record. So in this case, we are looking at record one, and we can see that she has been exposed from record number fifty, and from record number thirteen. If we click on the Exposed To tab, we can see this case has exposed to record number ninety nine. Now, if I were to click on record fifty, record fifty will automatically load for me, and I can follow up with this case – I can look at the various foods that she may have eaten, and take a look at some of the symptoms that she may have had.

We can also see that record fifty has been exposed from case ID number seven, and if I were to click on this icon [for record seven in the Linked Records area] it would take me to case ID seven. Now one

thing the Exposed From and Exposed To tabs let us do is create a directionality of exposure. So in this case record seven has been exposed from these two records, and is exposing to record number ninety nine. We can view all these exposure relationships in a social network analysis graph by clicking the View SNA Graph button. The red dot in the center represents the case we are currently looking at, record number one, and we can see that record one has been exposed from (based on the direction of the arrows pointing to record one) record thirteen and record fifty, and is exposing to record number ninety nine. If I wanted, I could double click on any one of these circles to take me directly to that record.

Some of you might be wondering, how exactly are these forms created in Epi Info? Let me go back to the Epi Info menu [goes to main screen and clicks on Create Forms button]. Let's create a new form. I'm going to create a new project, we'll call this MedicalRecords, and we'll give it a form name of Survey. You'll see that we have the option of creating this project as a Microsoft Access database [chooses Microsoft Access 2002-2003 from the Data Repository pulldown menu] or we can create it in Microsoft SQL Server. I do want to point out that you don't have to have Microsoft Access installed on your system to use the Microsoft Access option. Because I don't have SQL server available to me I'm going to keep this [Microsoft Access] as the default. I'm going to click Okay.

At this point, I can simply start adding fields. I'm going to right-click on the canvas, and let's select a New Field. I'm going to create the Last Name field as a text field [selects Text from the pulldown menu and creates a field named Last Name] and let's create a field for the patient's age, and we want that as numeric data, so let's create that as a Number field. [selects Number from the pulldown menu and types Age in the Question or Prompt: text box] If I wanted to specify a Range of values for this field [checks the Range box and enters zero in Lower box and ninety nine in Upper box] and if I also wanted to create a Pattern that the input has to match. Let's add another field for the patient's date of birth – that's a Date field. Just like with the Number field, I can set a range of Lower and Upper date values [month, day and year] and maybe I want to give this questionnaire a title, so I'll choose Label/Title and select a nice big font so that it will stand out.

I can move these fields around by clicking and dragging the prompt [label] on the canvas. I can move the input box independently of the prompt to the left side, right side. If I wanted to go back to the default, I could simply right-click on the prompt and select Default Prompt Align. I can select the field as a group and move them all as one. I also have various options for aligning fields. I can also undo and redo various actions by using the Undo and Redo buttons at the very top of the screen.

Templates

One of the things we've found is that many of our users are creating many different data entry forms, and in many of these forms, they are using the same fields over and over again. Such as Last Name, First Name, Age, Address, Date of Birth, so forth. What we've done in Epi Info Seven is added a feature called Templates. What Templates allow you to do is very quickly create a standard set of fields. If you notice, we have this Template option on the left side of the screen. I'm simply going to

expand the Fields under Templates - we have a couple of them here, Demographics, Diagnoses, Geo_Location, and Medical_Facility. I'm simply going to click and drag the Demographics template onto the canvas. In a couple of seconds, it has created a number of demographic fields for me. Now I can modify this if I want, for example, if I didn't want to collect any information on the patient's phone numbers, I could simply highlight the phone number fields and delete them. I could rearrange those fields if I didn't like their placement. It's also very easy for users to create their own templates. For example, if I wanted to create a template for this last set of fields, I could simply select those fields, right-click on the canvas, and select this option that says Save Selection as Template [at the bottom of the pulldown menu that appears]. I will call this Demo Template [In the Template Name field in the Add Template popup box], and Demo Template now shows up in the list of templates that is now available to me [under the Template Fields area on the left side of the screen].

That concludes the demonstration of Epi Info version Seven. I'd like to thank everyone for their time and interest, and if you have any questions, please send them to epiinfo@cdc.gov.