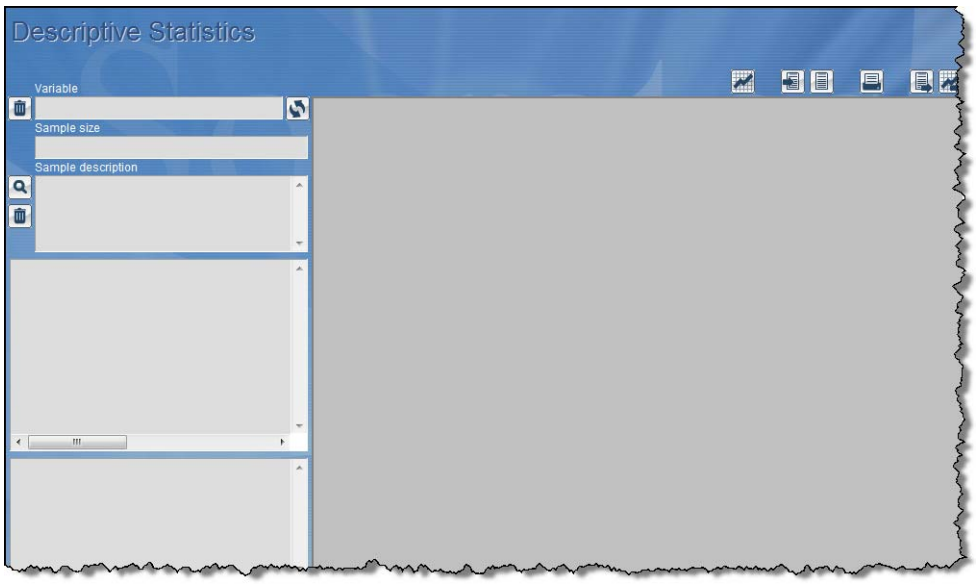
Statistics

This chapter will become important once you have collected enough data to start to generate some meaningful results. You can generate basic descriptive statistics on any of the fields in the database yourself using this feature which calculates statistics such as mean, median, %'s, standard deviations. For more sophisticated statistics you will need to export the raw data into a statistics package such as SPPS and enlist the help of someone familiar with statistics. Do this by using the Export function. You will need to be familiar with how the search function works in Socrates to be able to use this feature effectively, if you haven't already read that chapter it's a good idea to do this first and practise searching.

It's when you are ready to access and start generating results that the fun really starts. Here goes....

Go to the first screen and click on 'Statistics'. The following screen will appear. Click on the icon.

What you see next is a pretty boring looking empty screen.

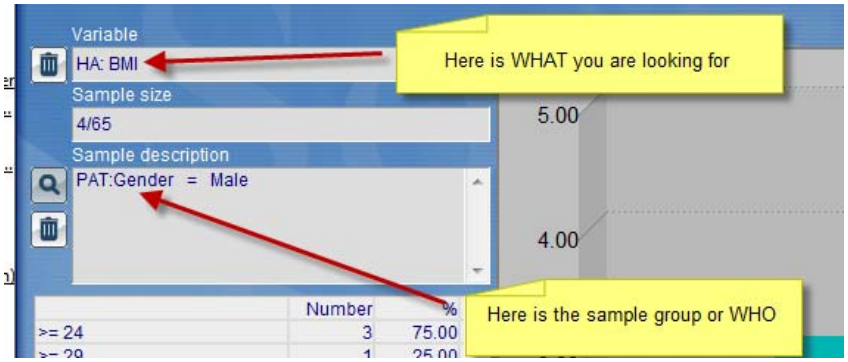


Before you get started you have to decide 2 main things:

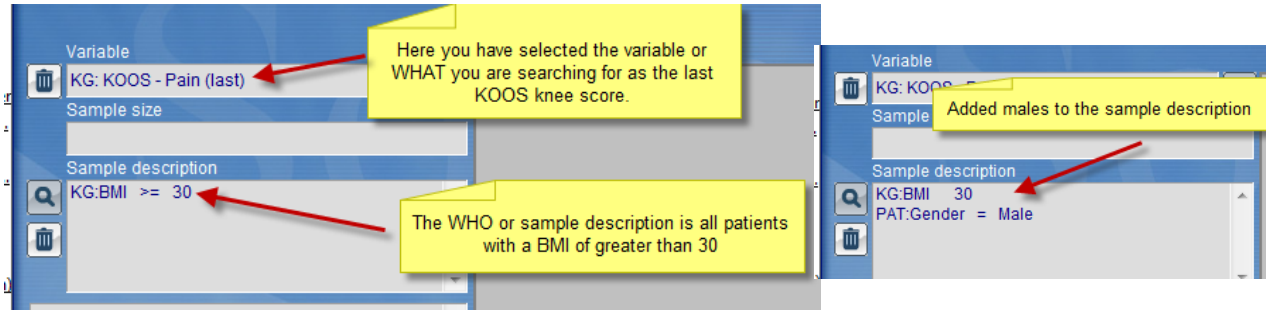
FIRST: **What are you looking to analyse?** A mean score, a summary of what you have done, a list of complications, the number of operations you have done etc etc. We have called this your VARIABLE.

THEN: **Who or which group do you want to know this about?** All records or a subset? Maybe compare males to females, primary cases to revisions, those with a BMI > 30 vs BMI < 30 etc etc. We have called the fields which would be displayed here your SAMPLE DESCRIPTION; it's a description of the sample of the data base you have chosen to analyse for the selected variable.

It's possible that the same variable could be used for both the variable and the sample description – what and who – for example BMI. If what you are looking for is the BMI on the whole group, then males vs. females you select BMI as your variable. The sample description could either be left empty and you could get the BMI for everyone, or you could select only Males, then females. In this case the **WHAT** you are looking for is the **BMI** the **WHO,** or the sample description, is either left empty for everyone or shows males then another search for females.



Then you might want to see if there is a difference in a KOOS knee score for a BMI over a certain value, say 30. In this case you would select the *score* you want as the variable, then in the sample description you would select *BMI* > 30, and you could also add male and then females and compare the differences. In this case the WHAT you are looking for is KOOS knee score and the WHO or the sample description is the BMI > 30 and then males/females.



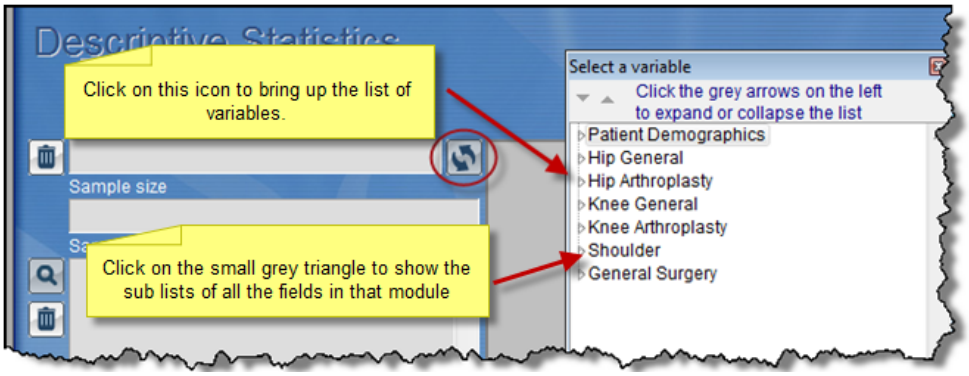
WHAT ARE YOU LOOKING TO ANALYSE?

In order to be able to use the stats function you need to first know what results/statistics you are looking for. The statistics are independent of patient names, they don't list any identifiable details, just analyse the data you have input. So before you go to use the variable window you need to have figured out just what variables you want to analyse. Some examples might be the:

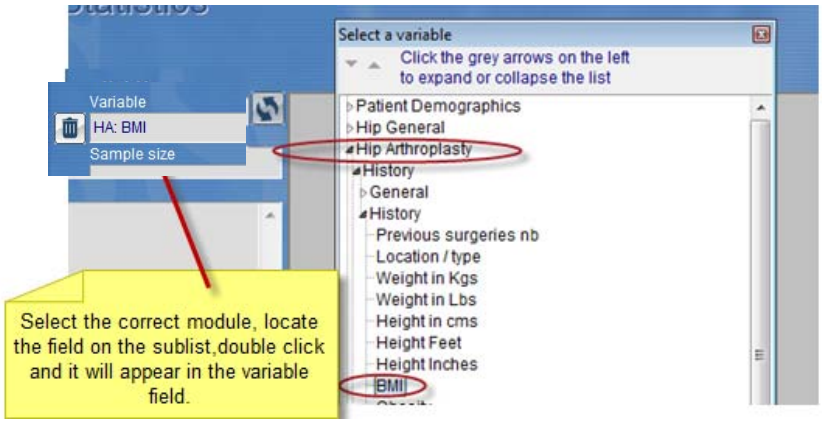
- Scores for pre op vs. the last evaluation, or an evaluation at a set time point, 3-6 months for comparison.
- Mean and medians of fields that are numeric values - number of previous surgeries, tourniquet time, leg length, time from injury to surgery etc. Remember you can also use the search function to do stats on a subgroup. For example if you wanted to look for the scores on patients who had equal or less than 2 previous surgeries you would use the search function to select this group after you had selected this stats Variable to analyse the scores.
- Complications – you can generate a report/graph of all your complications. You can select a list for all who had significant comorbidities and then a list for those who didn't, for example. The graph below displays complications for the whole dataset since the sample description window is empty.

SELECTING THE VARIABLE/S TO ANALYSE

Once you have decided what you want to analyse you need to select this variable to generate the reports. To do this, click on the squiggly icon shown below. This will then bring up a list containing all the variables in the database which are separated into modules to make the lists manageable. They are also listed inside this as sub lists, divided into what you see on the screen, history, surgery, follow up etc. The Search chapter tells you how to use this function in more detail.

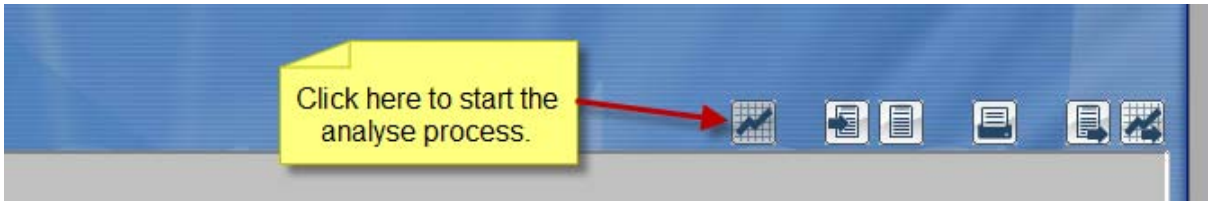


You then need to choose the module you want to generate your statistics on. In the next example we chose the Hip Arthroplasty module (HA) and then the BMI. Double click on the one you want and it will be displayed in the variable window. The variables are sorted into the same order and groups that they are displayed on the screens within the program. If you wanted to find the BMI in the example you would look in the history section since this appears on the history screen on the program. The sample description window in this example is empty so you would get the BMI's for everyone in the HA (Hip Arthroplasty) module once you clicked on the icon to tell the program to do the calculation.



The Analyse icon

Now that you have decided what you want to analyse, and for who or which group or sample, you would click on the ANALYSE icon.

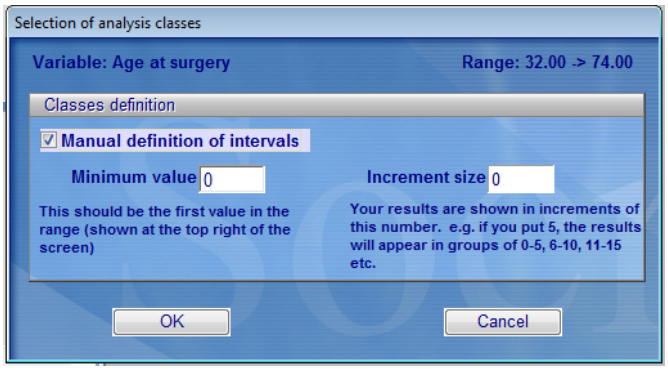


Once you have chosen 'Analyse', the program will either produce the results or ask for more information.

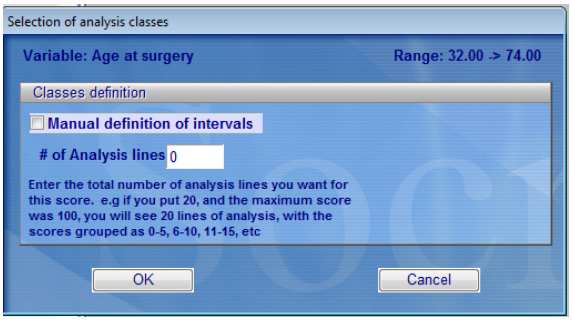
Numeric Values

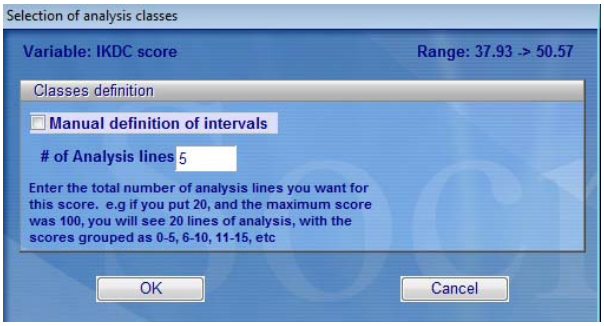
If the variable you are looking for is a numeric value of any sort you will need to select the increments you want the number/s displayed or grouped into. One of 2 methods can be chosen to display and select this.

- This is normally the default method. The top right numbers display the range in this group – the lowest is 32, the highest 74 in this group above. You can then select the *minimum value* which would normally be the lowest number in the range and then select the increment size, which is how you want the group to be clustered.

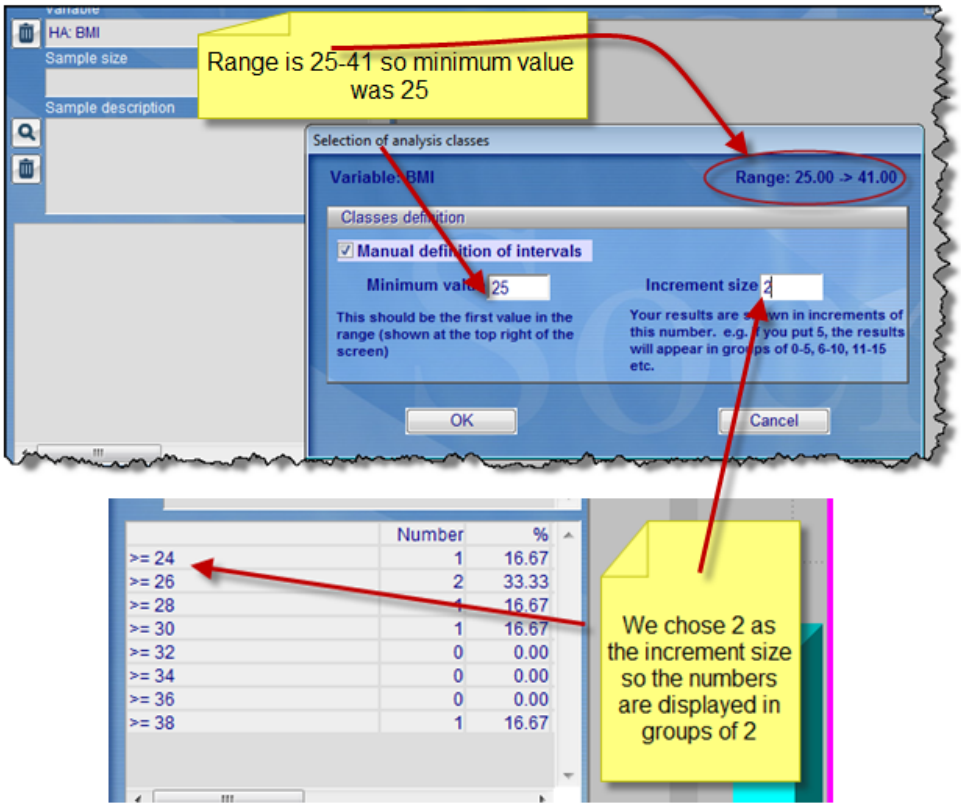


- Or you can just uncheck the 'manual definition of intervals' and write in a number that corresponds to how you want them grouped. It will automatically start from the minimum number. The example on the right has selected the score to be grouped in 5's.

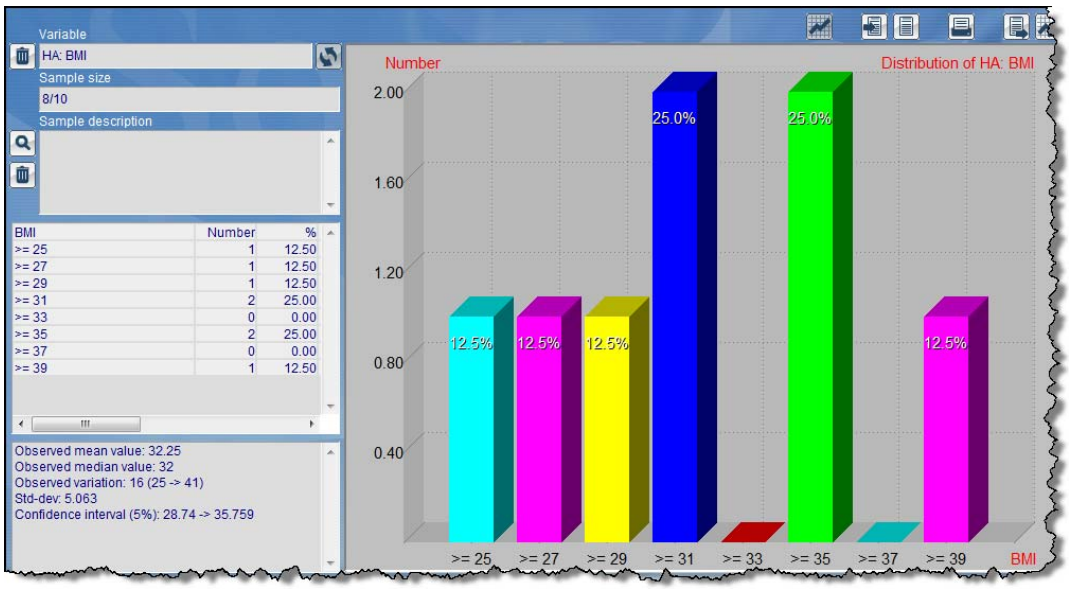




Selecting minimum value and increment size. If the range is small, as in the next example 25-41, you might want to select to group the results in 2, or 5. If the range is wide say 5-2075 you would want to select a larger increment size or you would end up with a very long list. In this case you might want to group them in clusters of 100, or even larger.



And, here's the result. It's the same no matter which method you chose. As there's nothing in the Sample description window it's given us the BMI for all surgeries in the HA module. You can see the mean, median, standard deviation etc. and the number that were analysable – 8 of the 10 records had the BMI recorded.

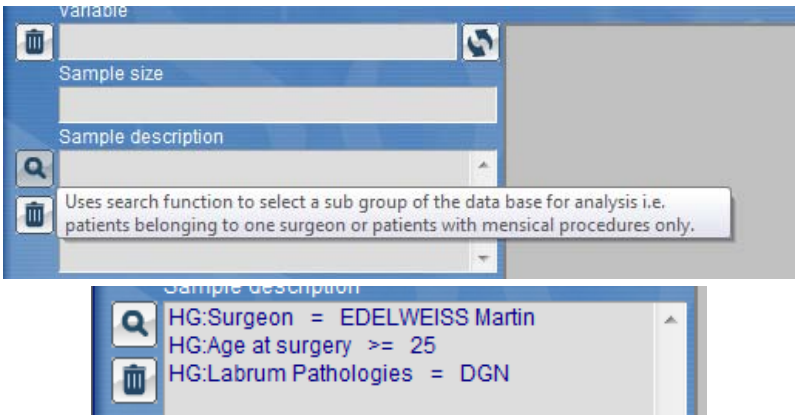


WHICH SAMPLE GROUP OR SUBSET DO YOU WANT TO ANALYSE?

Now that you've decided what you want to analyse, which groups of patients or surgeries do you want to do your stats on. Is it all the surgeries in the module or a sub set of them? You can generate statistics for the whole database or any subsets of the database. Under the 'Sample Description' window you can use the search function in the normal way to search for a sub-set.

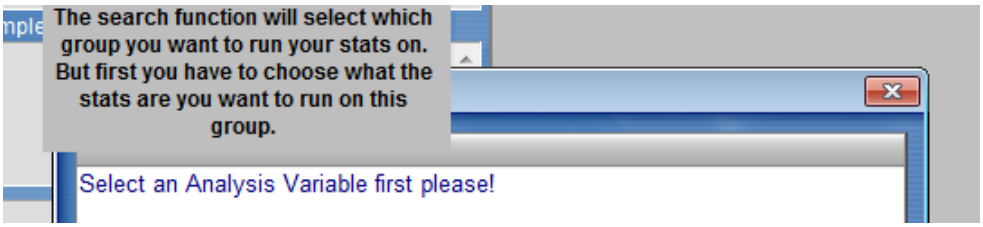
Note that the program runs statistics on surgeries, not patients. So if a patient has had more than one surgery they will show up in the stats twice. Don't forget that you can use this to select out those you don't want, so you could choose just primaries, or not include those which are revisions. Every variable on the database can be included or excluded.

Remember that when the window labelled 'Sample description' is empty the statistics will be generated on the entire database. If you want to look at a subset of the database, i.e. a particular surgeon, surgery group, males only etc. use the search icon on the left - the sample description to choose just the group you want to analyse.

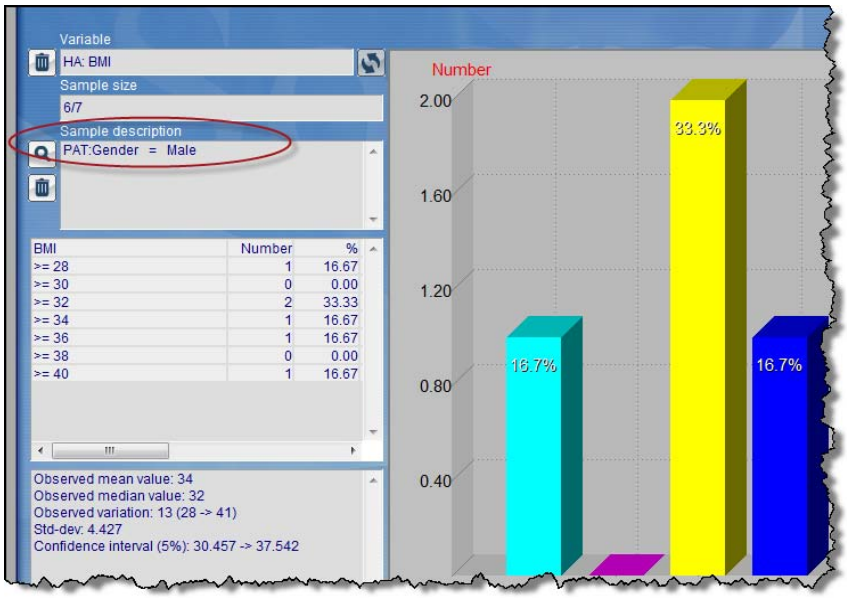


You can choose more than one group at a time as shown above, i.e. Surgeon is Edelweiss, age = > 25, labral pathology is degenerative. All your statistics will be generated on this group until you go back and delete these or change to select another group. For more information on how to use the search function go to the chapter on How to Search. After you have selected the group you want, or left it blank for the entire data file, you can start your analysis.

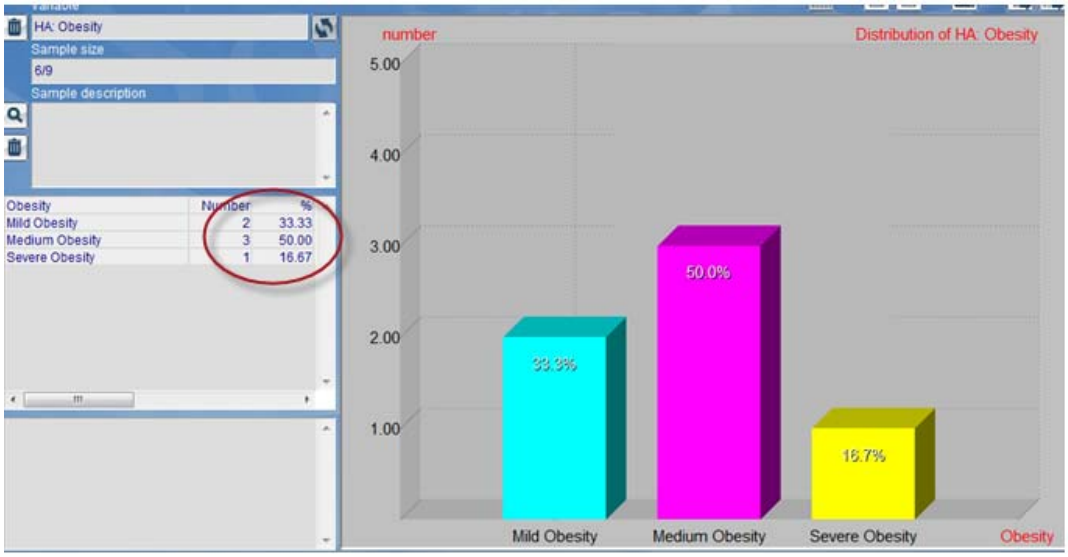
But remember that before you select whether to run the stats on all or part of your surgeries you first have to choose **what** you want to analyse using the top squiggly arrows. If you haven't done this and try to select a sub group you will get this message.

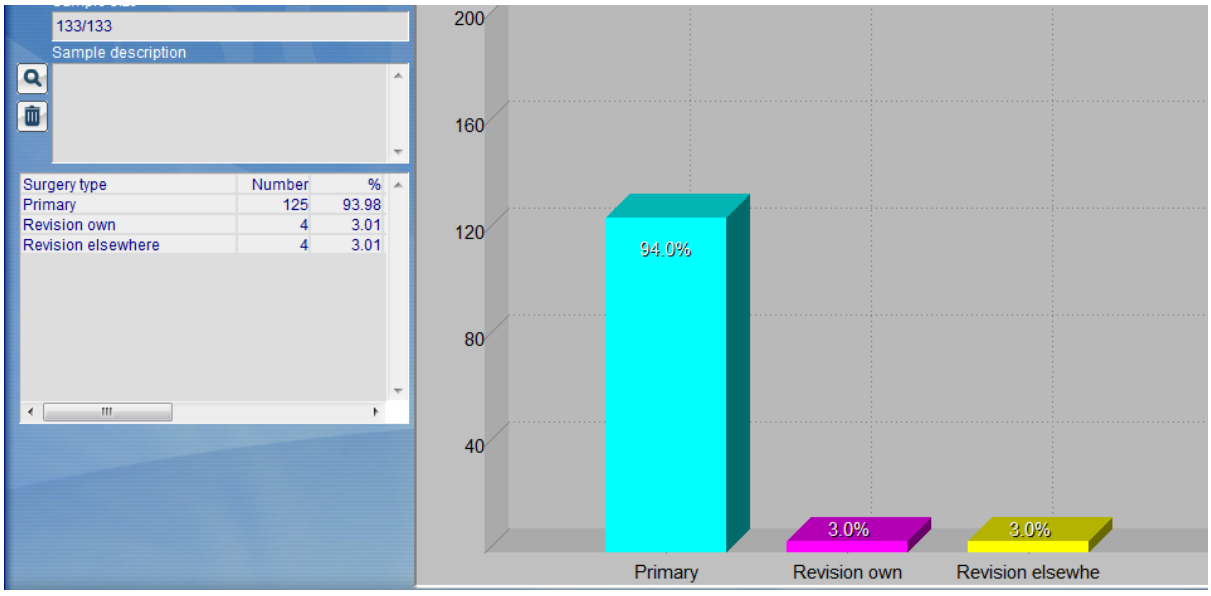


Below we selected Males only and now have the mean and median BMI on this sub set.



Values which are not numeric will be displayed as a group with the % in each group as below.

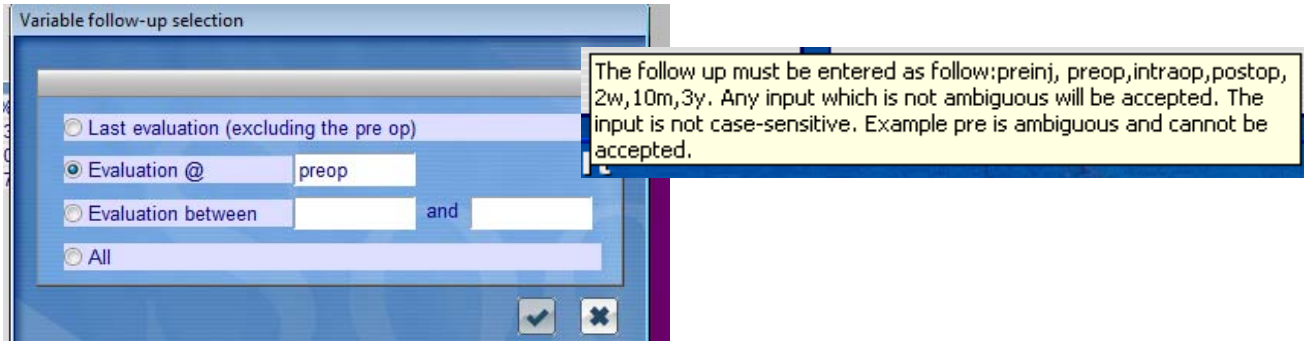




SELECTING DELAYS FOR EVALUATIONS

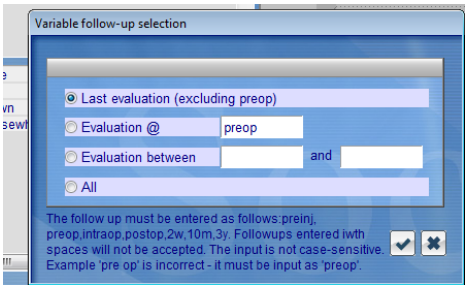
When analysing scores/evaluations, you will be asked to select the follow-up delays you want, since it usually doesn't make sense to mix pre and post op scores.

After selecting the score you want, the window below will be displayed. You must select one of the options. Options are Preinj (preinjury) preop, postop (less than 1 week) or any numeric value followed by either w (weeks) m (months) or y (years). Thus if you wanted between 12 months and 2 years you would select 12m and 2y as the time points.

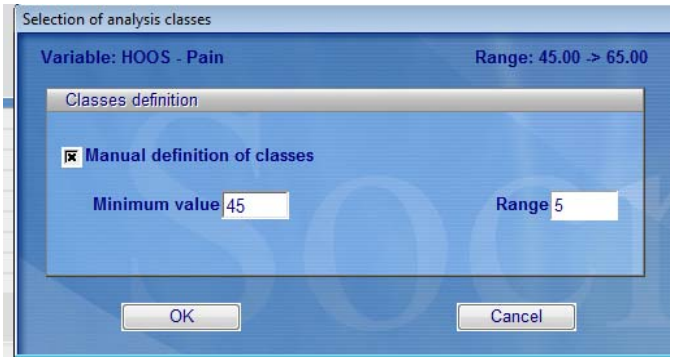


The latest score in that follow-up period for each surgery in the group you have selected will be analysed.

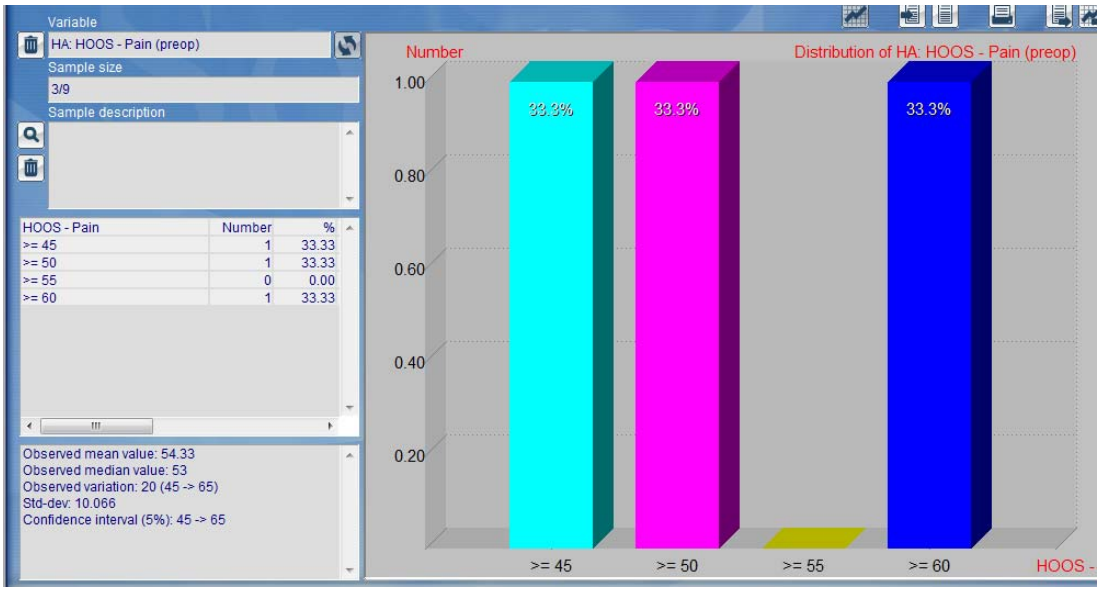
Below we are in Hip Arthroplasty, and have selected the Surgery follow-up and the HOOS score. Inside the HOOS all the questions are displayed and then the scores. We have selected HOOS pain score.



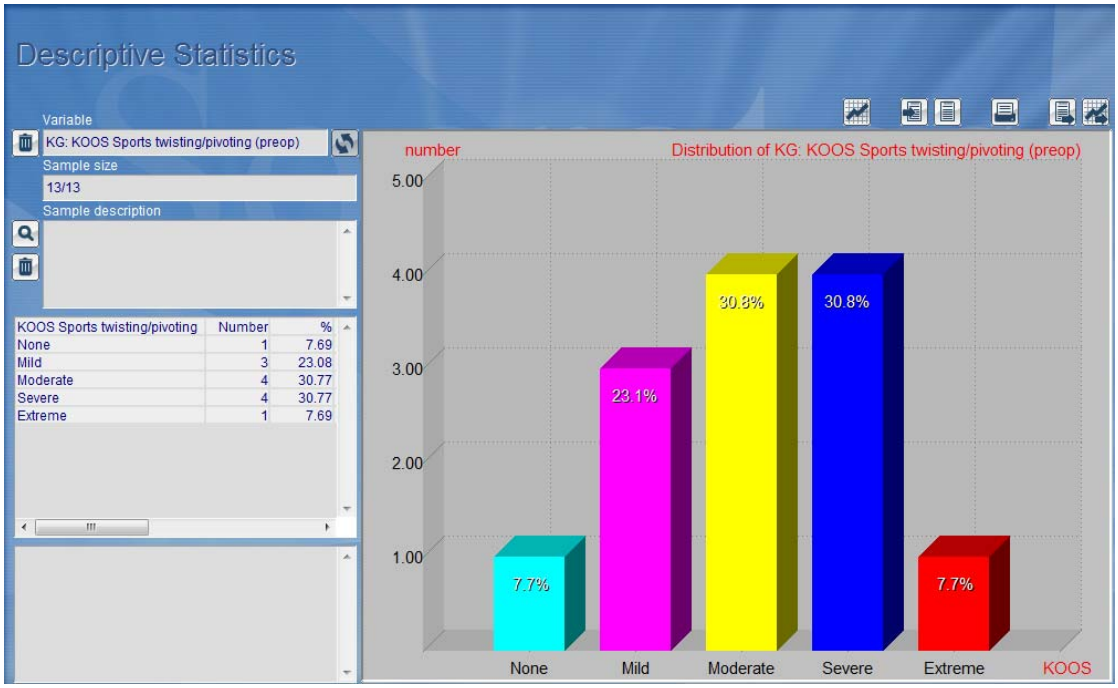
After selecting the time point you want to analyse, check the tick and the window below will appear. The top right hand corner displays the range, in this case it's between 45 and 65. The minimum value can then be entered as 45, the range is asking for how you want the numbers grouped, i.e. groups of 5, 10, etc. If the range is wide, say 1-100 you wouldn't choose to display them in 1's since the list would be very long. The next example shows them in groups of 5.

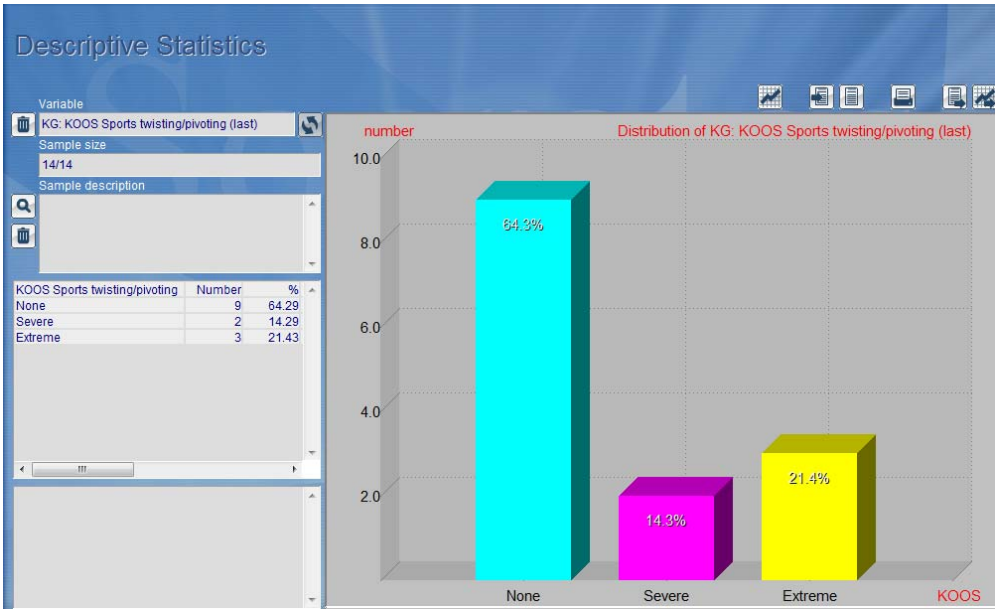


Select the range, click OK, and the statistics and graph will be displayed.



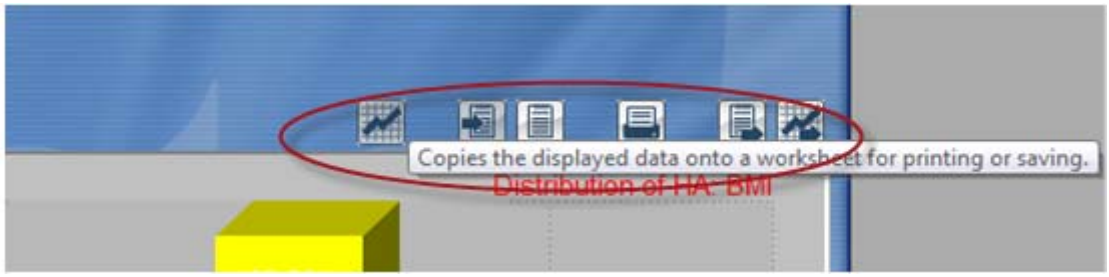
In addition to analysing the scores it is possible to choose to have each question displayed so you can see if there is a difference between some of the functions/questions that make up the score. The example below shows the question about how much difficulty patients had with twisting and pivoting pre op, and the next graph at their last visit.





What to do now?

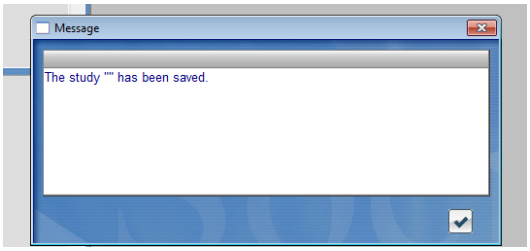
There are several ways of dealing with these data from here. Normally you would be analysing more than one parameter and will want to save the data onto one worksheet. The icons along the top enable you to save, export etc.



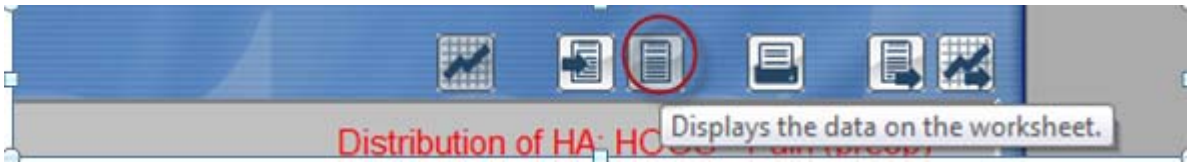
Exporting analysed data for graphs, presentations

There are 2 ways to do this. The first is to save the analysed data with the description of the data analysed to a worksheet. To do this click on the icon after the graph.

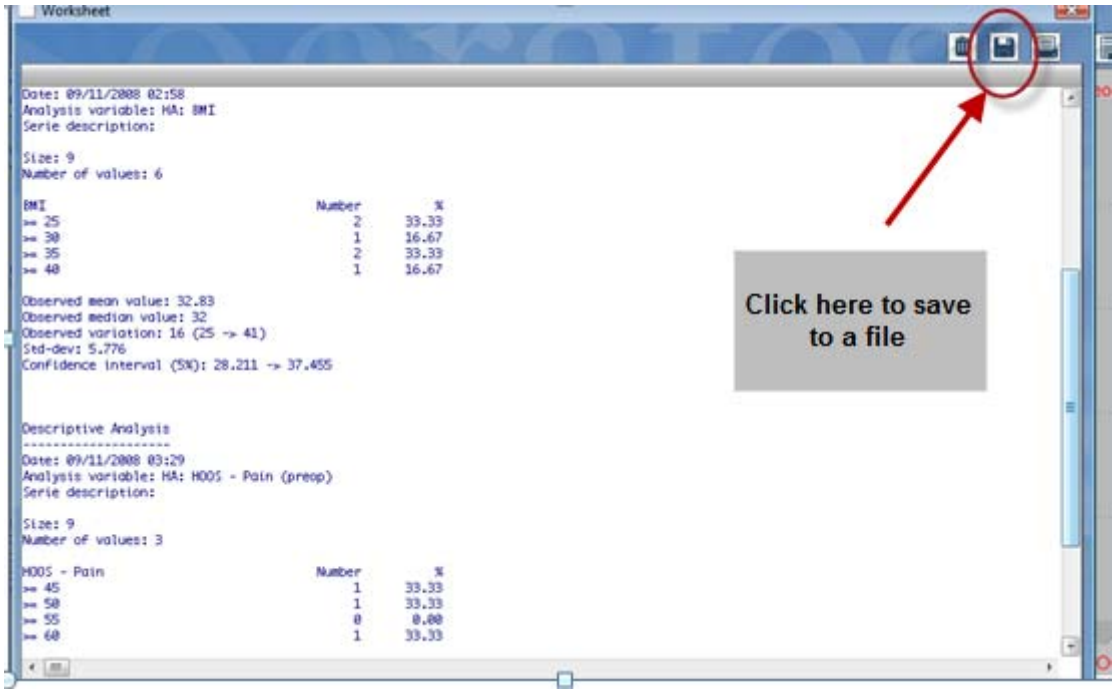
You will see this message telling you the data has been saved to the worksheet.



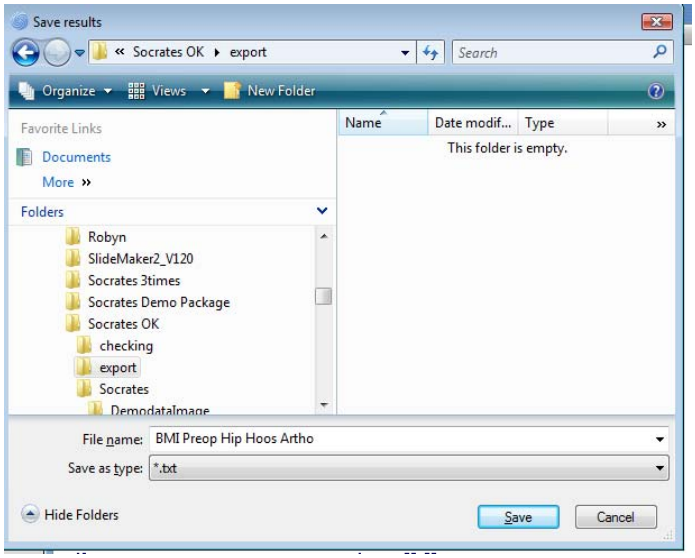
This can be done with multiple data sets. When you have completed all the fields you want to analyse you can then export the worksheet to a text file. To do this click on click on 'displays the data on the worksheet' icon (circled below).



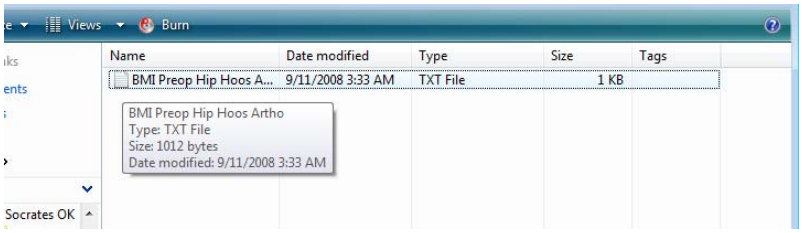
The worksheet will display as the next example. You can either print this out or save it.



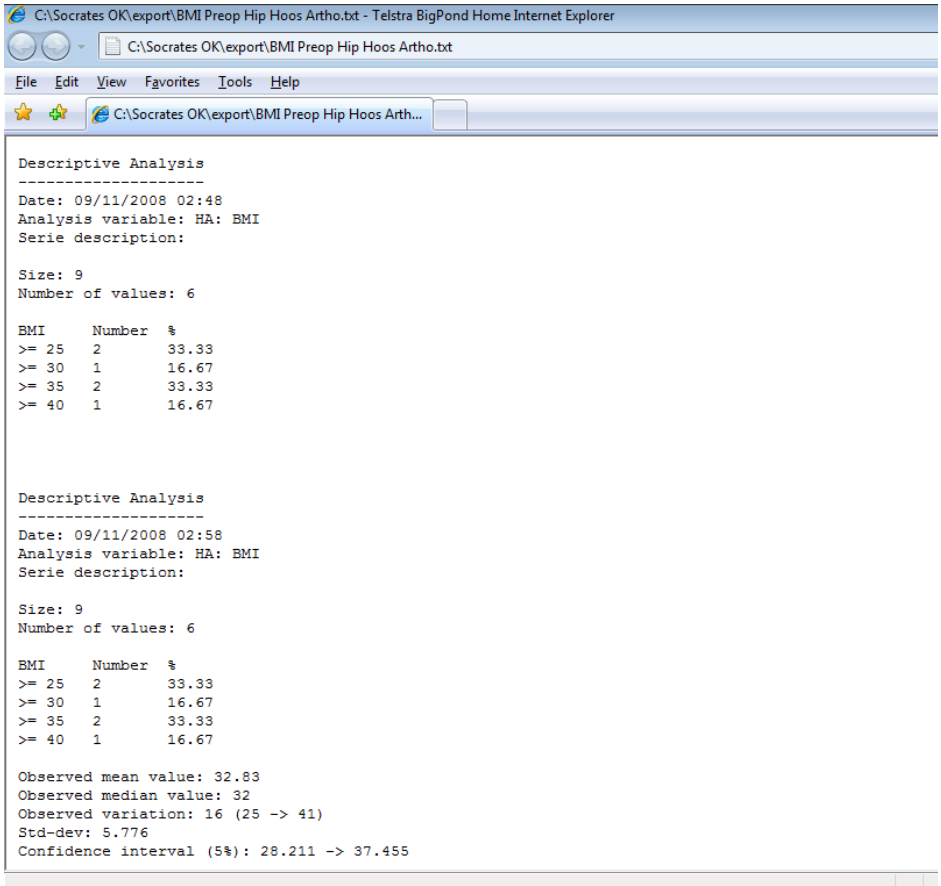
If you elect to Save, choose a name and location for your file.



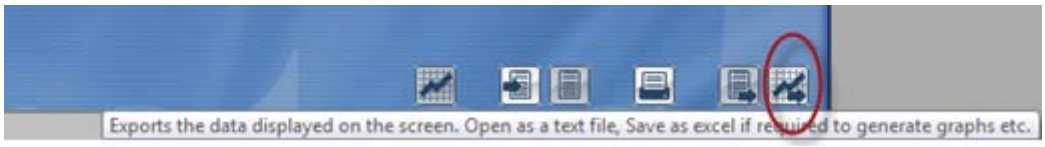
It will save as a text file which can be exported to any program supporting this format. To open, locate the file and open in the usual manner.



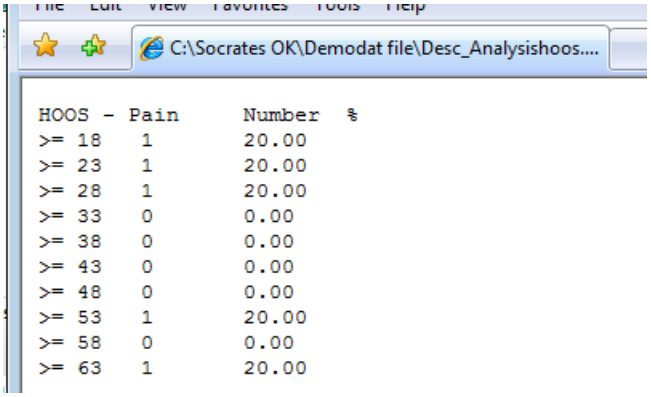
It will be displayed as below.



If you want to export just the statistics minus the text, go to this icon.

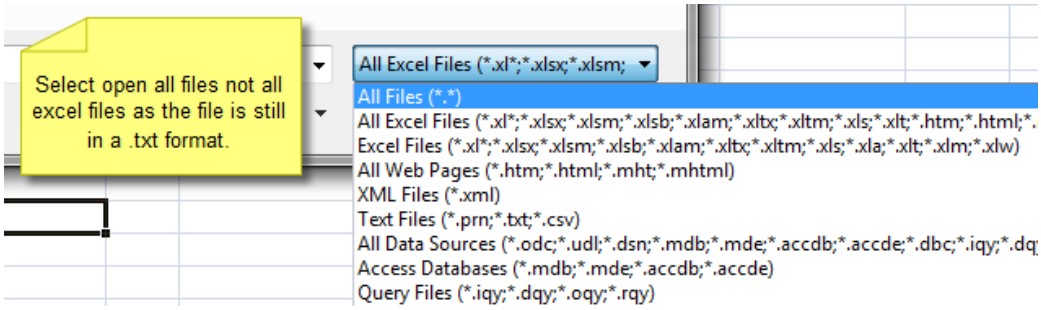


Save to a location you choose, the data will be saved in a text file. Only the actual statistics will be exported as below.

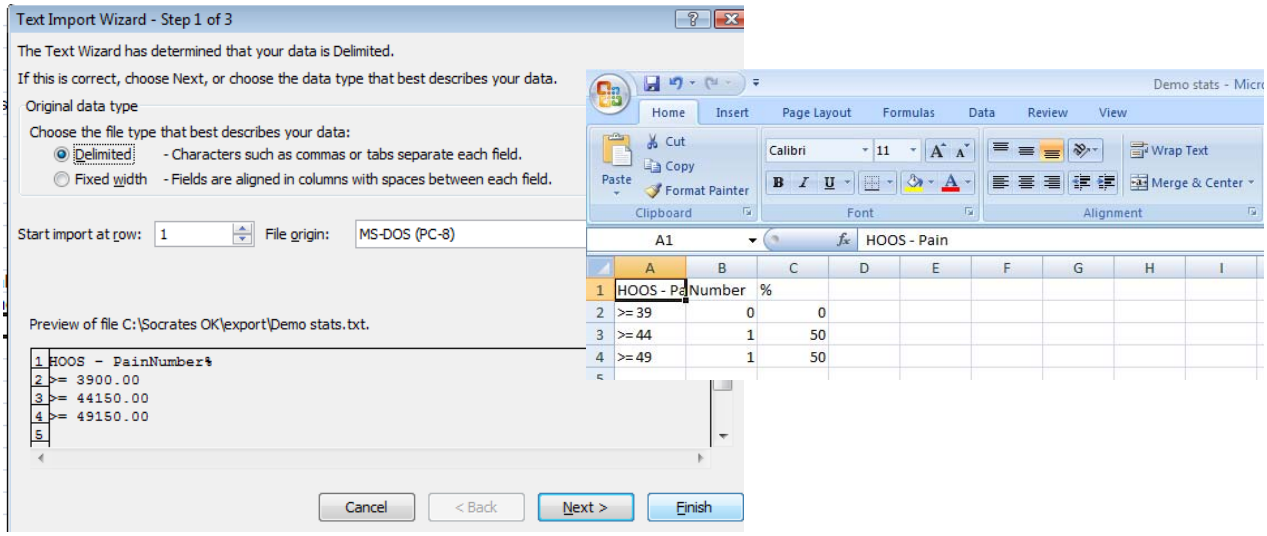


How to open text files into Excel

To open either of the text files into excel - Open Excel – select ***Open all file types*** (since it is still a text file).



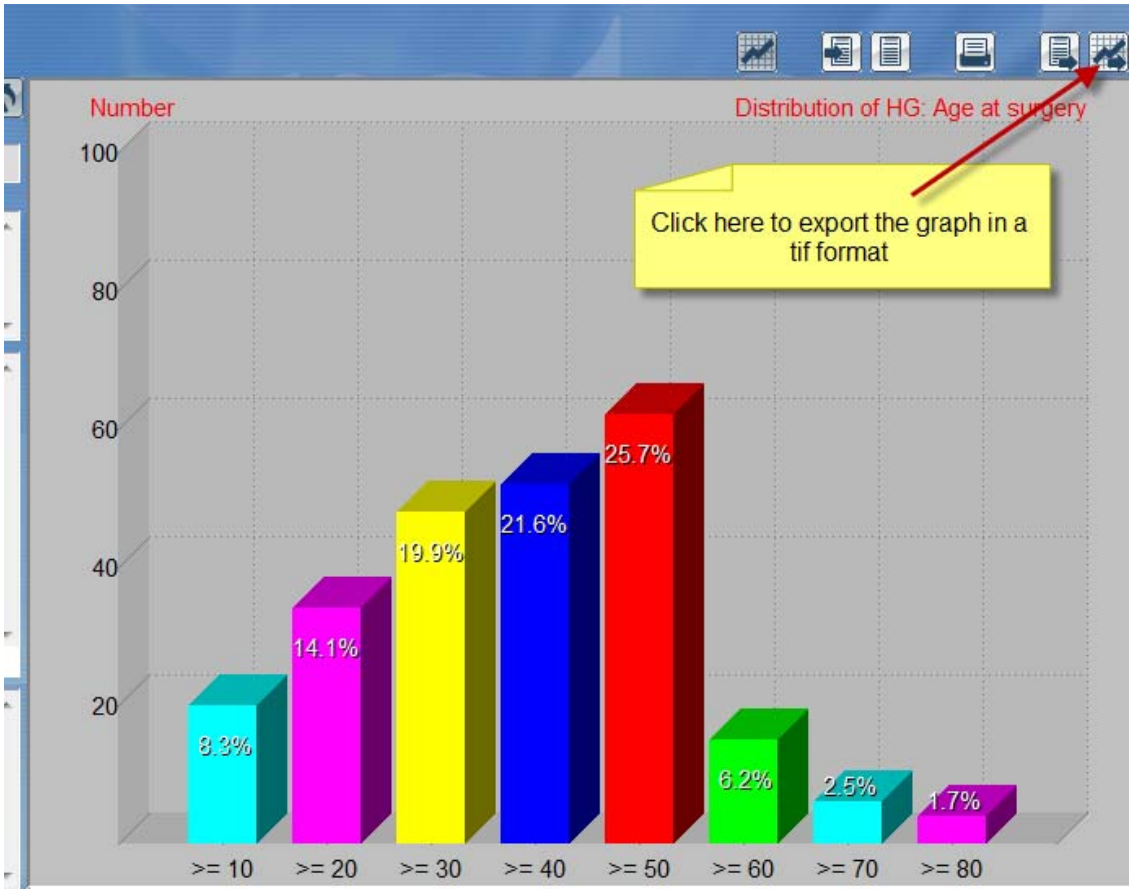
It will appear as a tab delimited text file as below, click on Finish and it will appear in normal Excel format. Save as an Excel file if you want to keep it in the Excel format.

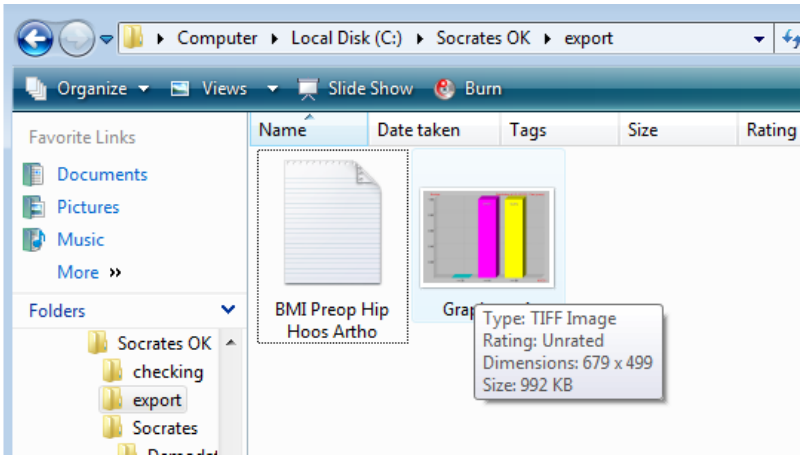


This data can then be put into a graph using Excel.

GRAPHS

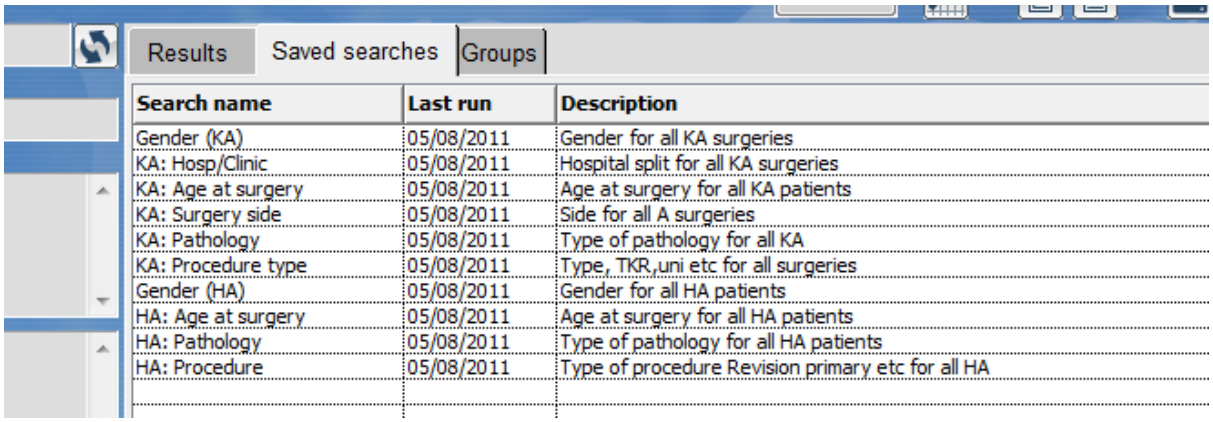
The graphs from the screen can be exported as a TIFF file and then imported into a PowerPoint or Word document. However, these are exported as an image thus can't be modified in any way. If they need to be relabelled or modified, the data should be exported to Excel as described above and the graphs generated from within Excel.





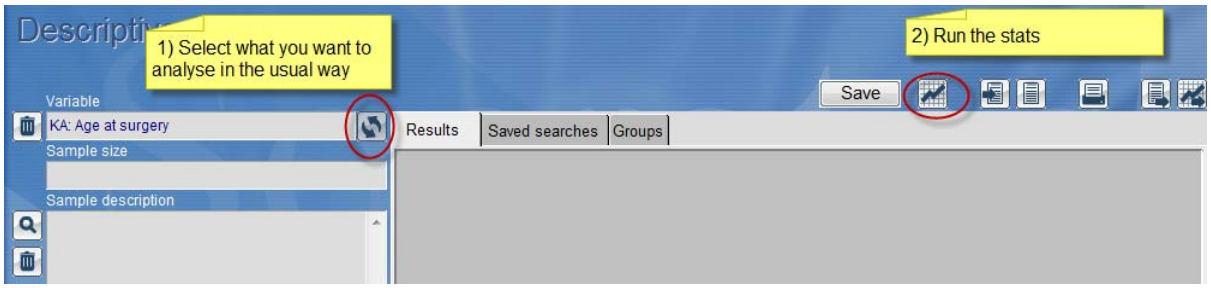
Saving Statistics

It's possible to save the statistics that you run routinely. For example every 6 months you might want to see the average age of patients, gender breakdown, and preop and latest knee scores for a sub set. You can save these as either individual searches/stats as well as group them together and run them all in one search without having to reselect each one each time. The example shows a list of statistics searches that have been saved.

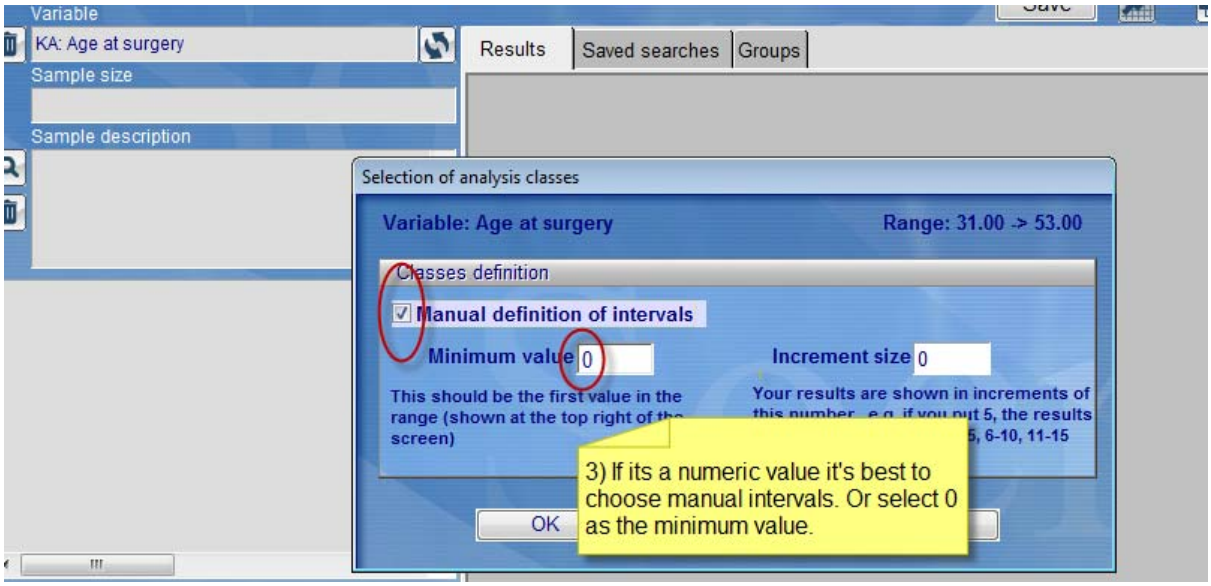


Saving your statistic searches

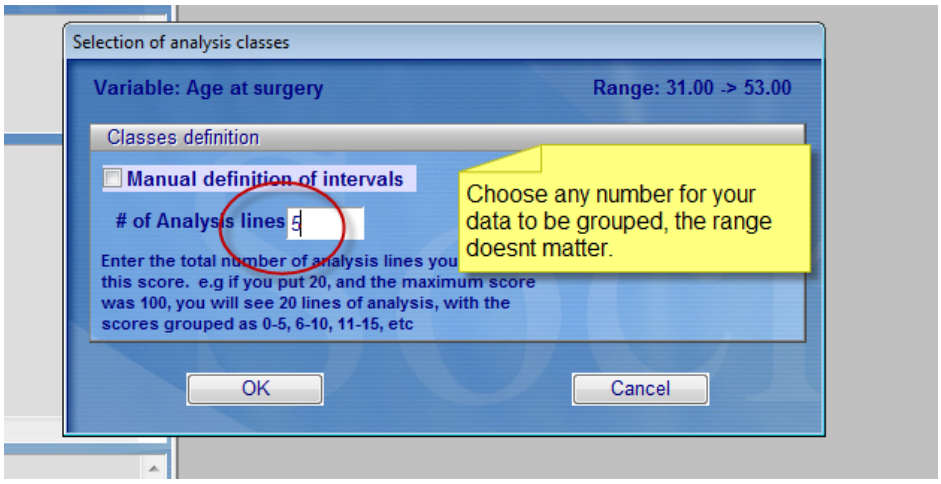
Use the normal method to run the stats for whatever you want. Then select the graph icon to start the analysis.



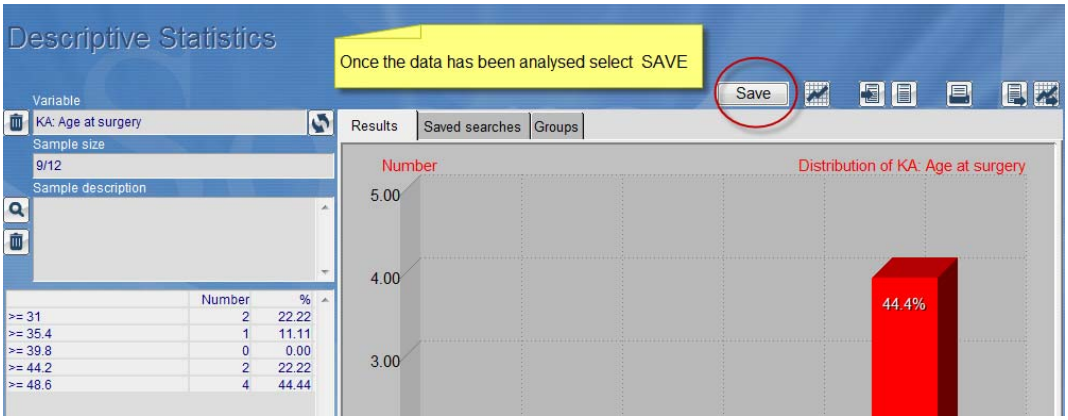
If the data you want is a score, or a numeric value its best to choose either manual intervals or select 0 as the minimum value. If you select the minimum as the lowest number in the range for this search it may not be the same range next time you run it. Thus you have saved these parameters you risk excluding some records from the search.



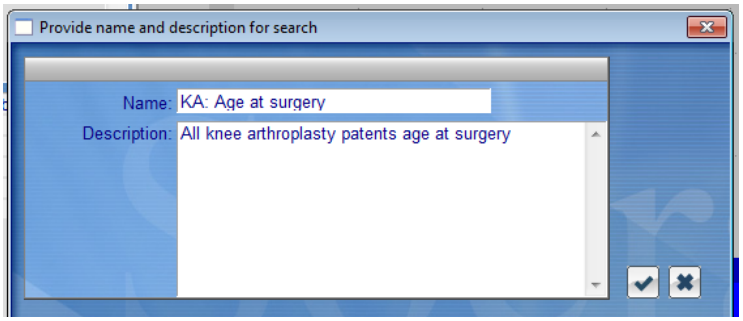
If you select manual definition of the intervals you can elect to have the values displayed in whatever groups you choose, and it will display the values in groups of that number and will start at the lower end of the range.



After your search has run, click on the **Save** icon.

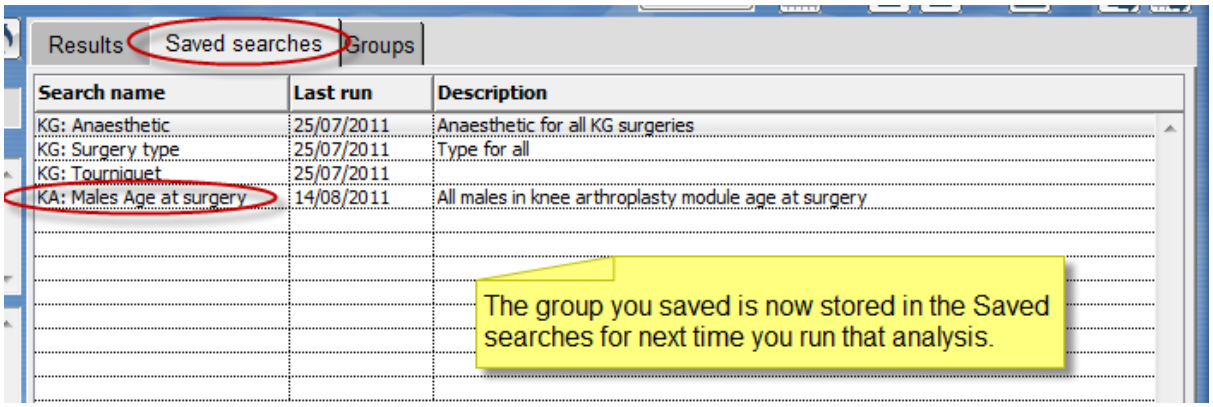


You will then need to name your search, you can leave the default name, change it, and add some additional details below to ensure you know which group you are selecting next time. Select the tick when you are happy with the name.

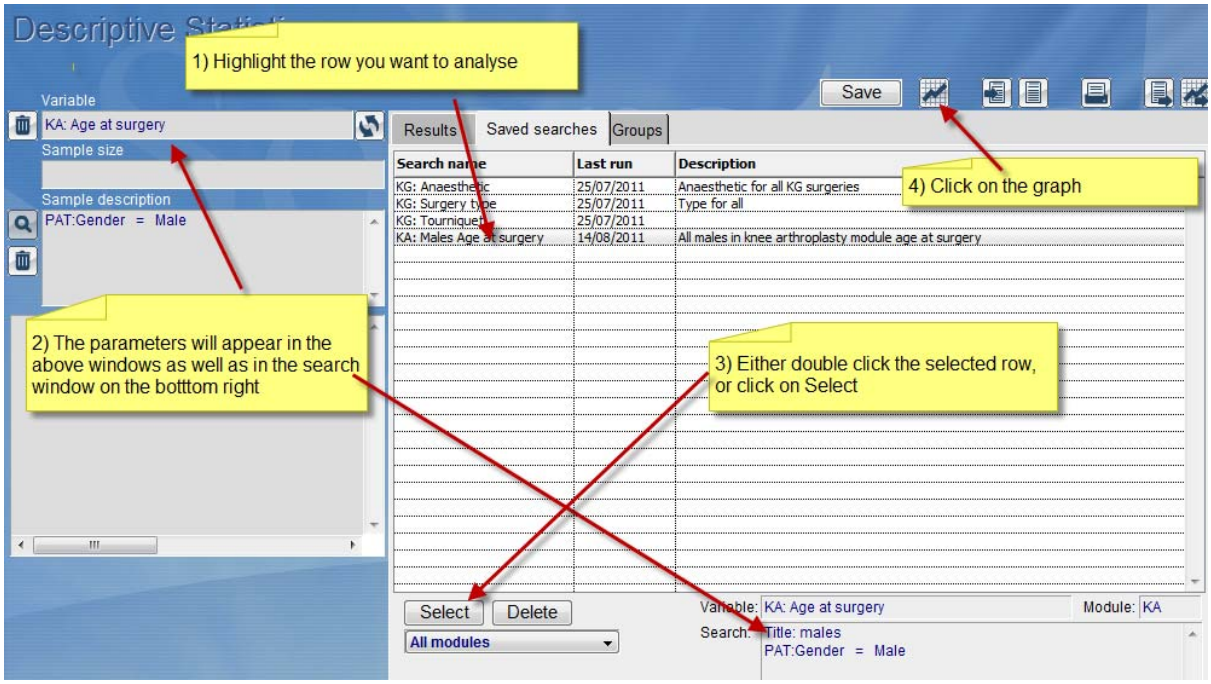


If you have selected a sub set as the next example (males) make sure you include that in the name.

Once you select the tick, it's saved for next time.



To run the search next time go to the Saved Searches tab and select the row where your search is. Either double click or click on the select icon, then the graph as per usual. The search will run as it would had you gone through the process of selecting the parameters manually.

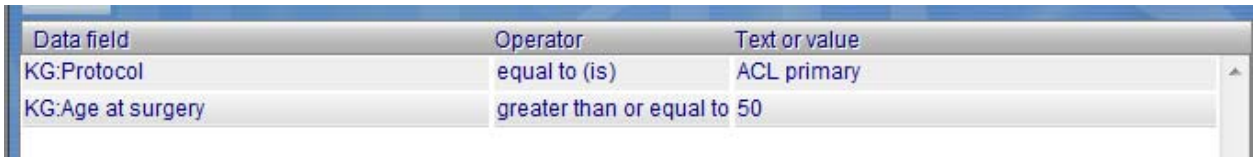


OK, so now we know how to save your searches, but you would rarely run an isolated one off search. Its possible to group them so that you can run a batch of stats with one click and save the results to a worksheet.

Saved Searches

It's important to understand when you are saving your statistics that if you are using saved searches any changes made to the search after you have saved these will affect the saved statistics. For example, you saved a group of statistics using a saved search for your all surgeries in the *Routine ACL protocol*.

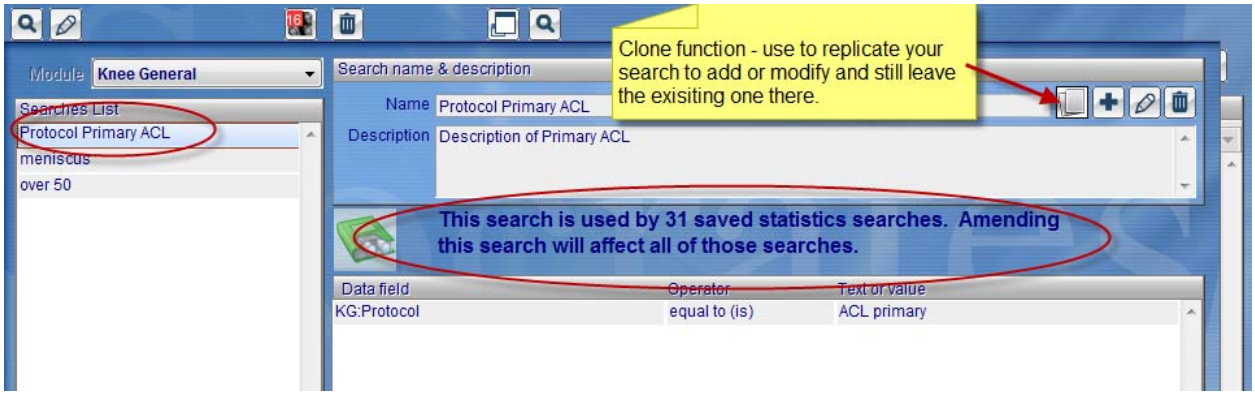
Then later you went into the search and modified the search to find a list of all patients over 50 in this protocol. So this search was now for all patients => 50 in the *Routine ACL protocol* . So this is now what your saved search is.



If you leave it like this when you run all your saved statistics it will now be looking for all patients in this protocol over 50 rather than the original of just all patients in the *Routine ACL protocol*.

To avoid this either create a new search with the addtional parameters in it, or delete the extra parameters after you have run your search. You can use the clone function to the left of the Add icon to replicate your search, make the changes here for your new search and still leave the existing one the same. Thus the statistics you saved will still be attached to the original search.

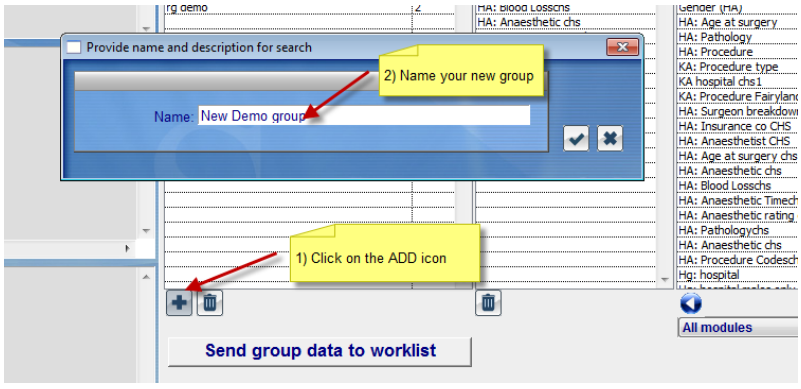
When you highlight the search you will see a message telling you the number of saved statistics searches which use the search.



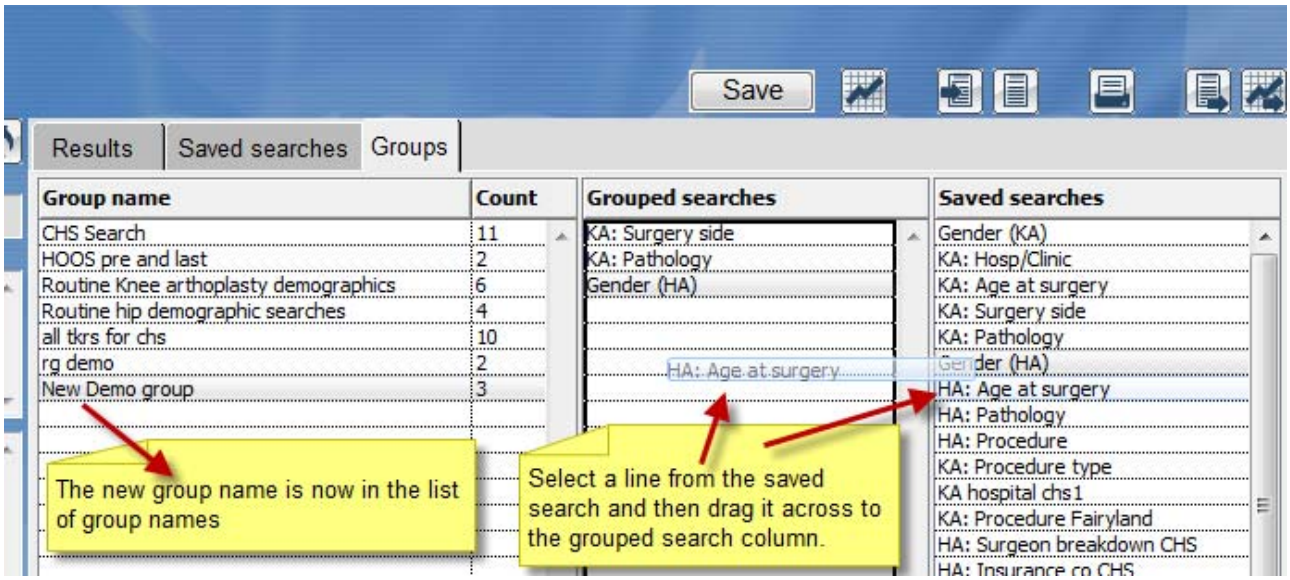
Saving Groups - Once you have saved your individual searches as described above, the saved searches can then be put into a group and the whole group can be run in one click. In the next example we have a group called *Routine hip demographic searches* which will run the Gender, age at surgery, pathology and type of surgery with just one click.

Saving the group

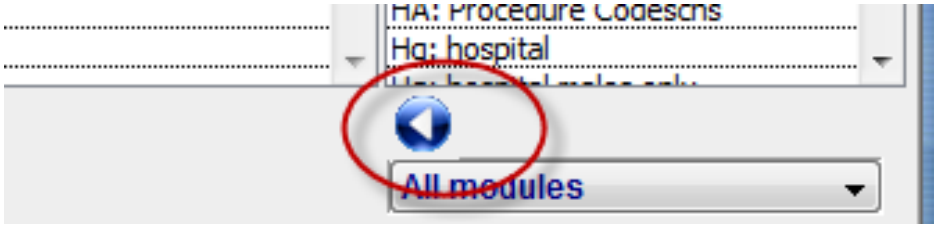
To save anything into a group you first need to have saved the individual searches. Once this is done select the ADD icon, then name your group.



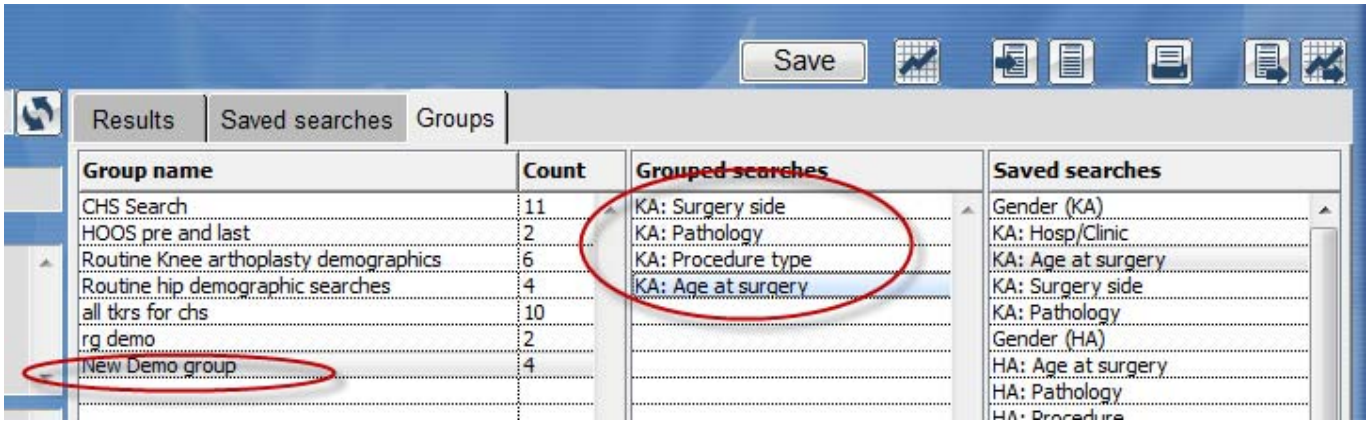
Once the group is named highlight it, then go to the last column and drag the searches that you want include in your group across to the middle column.



You can also highlght the search you want then use the back arrow to add it to the list in your group.

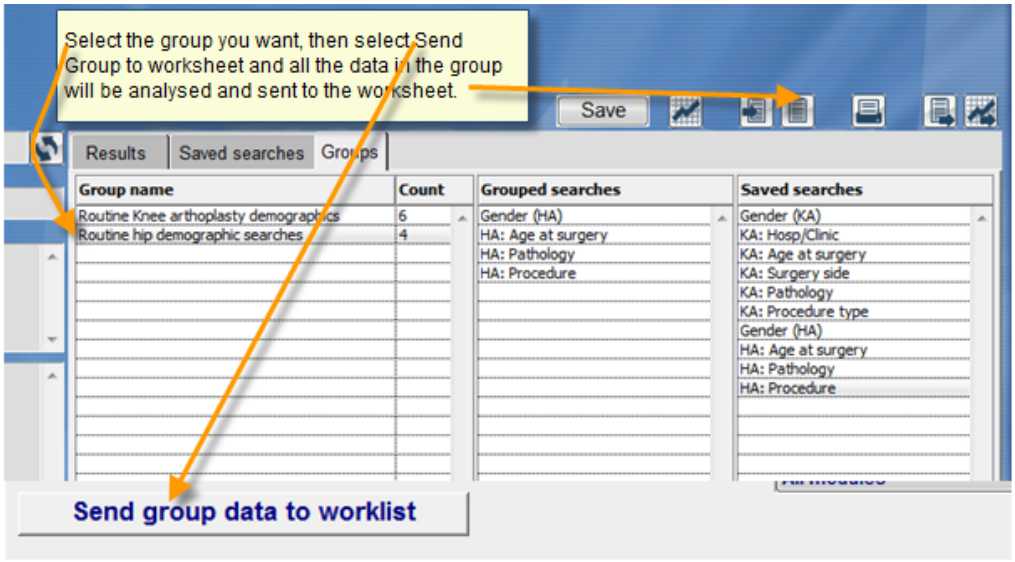


In the next example there are 4 searches in the New Demo Group.

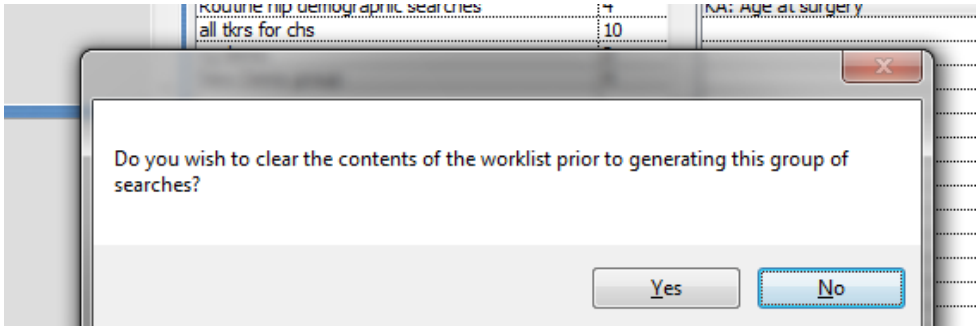


Running the Group Search

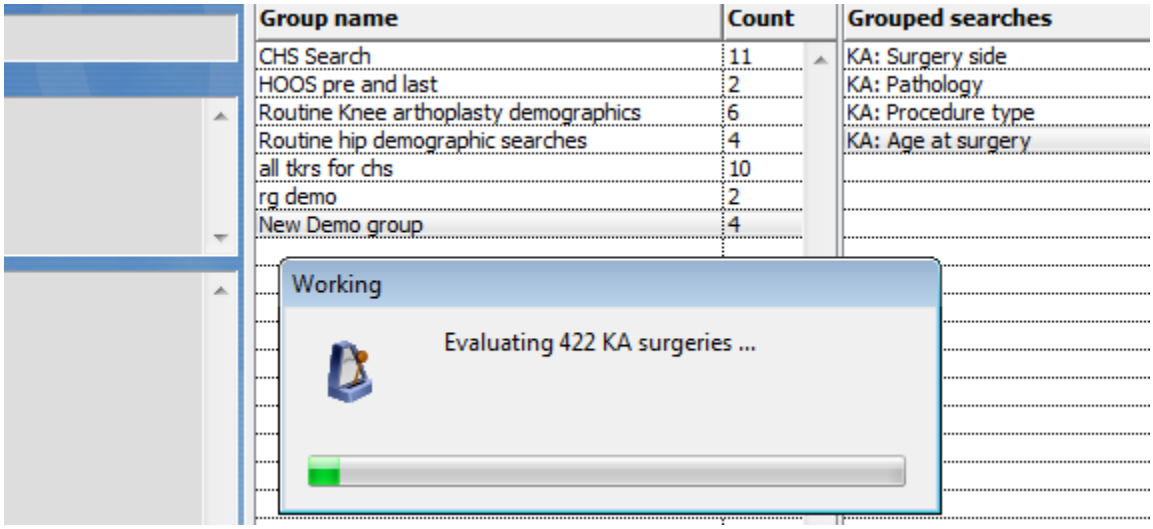
Highlight the group name you want then go to Send group data to Worklist.



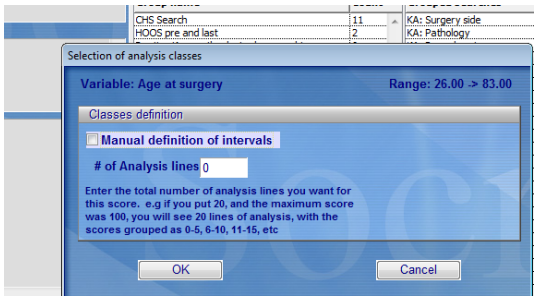
If there is already data on the worksheet you will see this message. Either select yes, or add your new group to the existing worksheet.



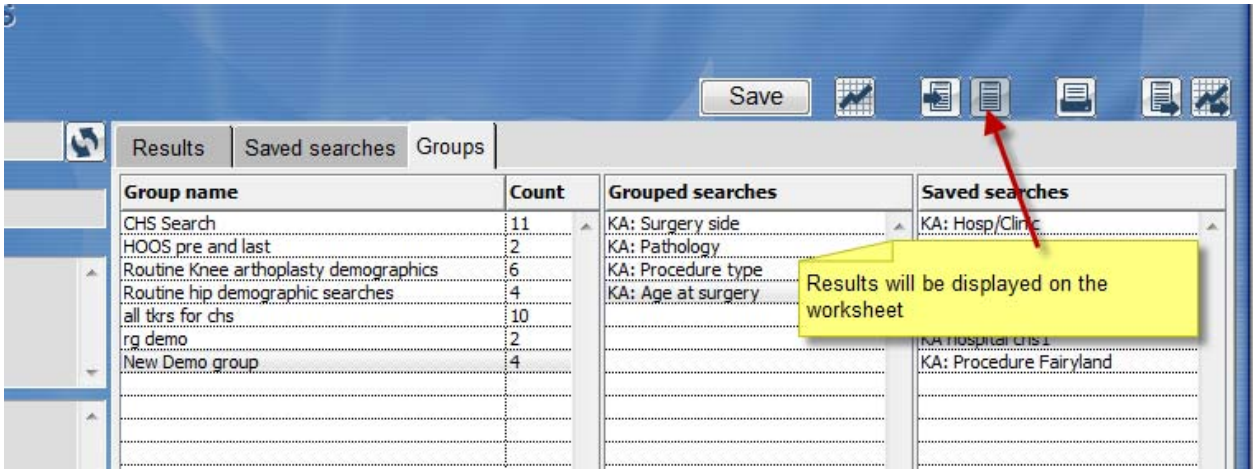
Once the group search starts you will see each search running through analysis process.



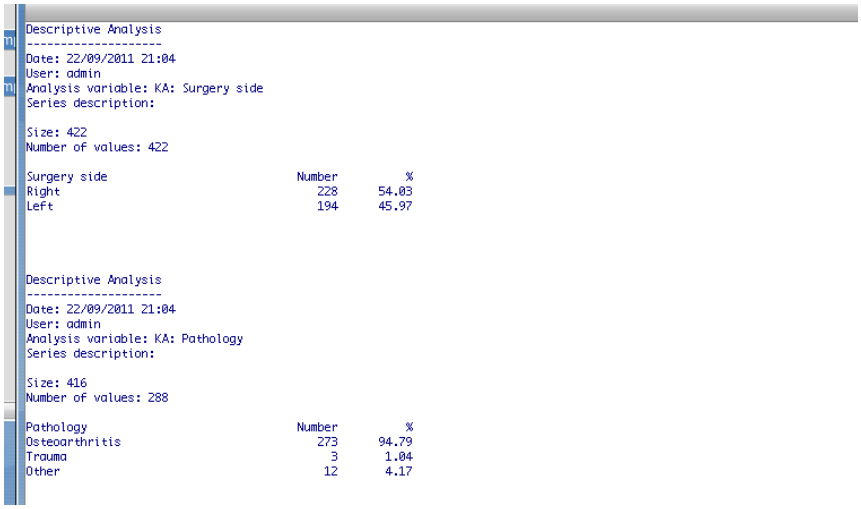
If some of the searches are numeric values the search will stop and ask for the range or interval which you will need to insert.



When its completed this message will be displayed.



The worksheet will display as below, and can then be opened in Excel.



Descriptive Analysis

Date: 22/09/2011 21:04
User: admin
Analysis variable: KA: Surgery side
Series description:

Size: 422
Number of values: 422

Surgery side	Number	%
Right	228	54.03
Left	194	45.97

Descriptive Analysis

Date: 22/09/2011 21:04
User: admin
Analysis variable: KA: Pathology
Series description:

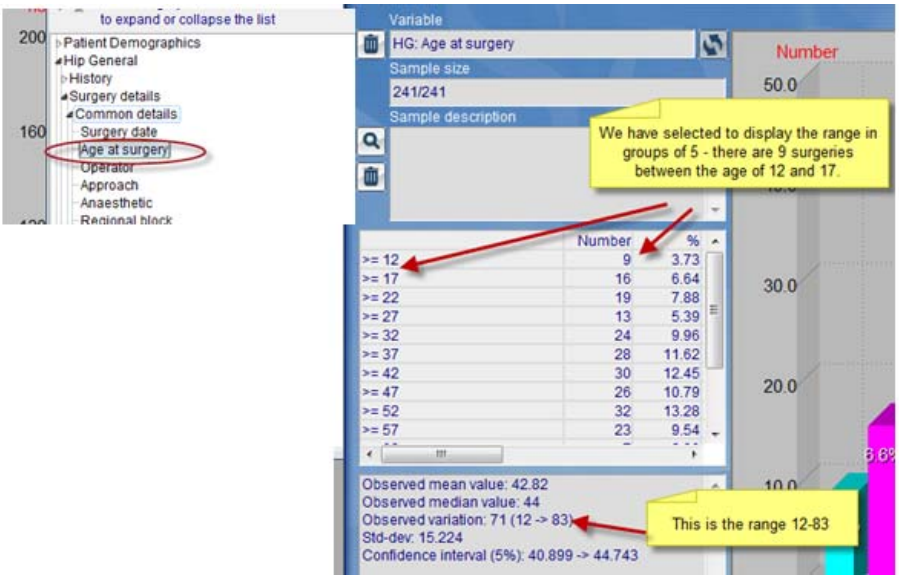
Size: 416
Number of values: 288

Pathology	Number	%
Osteoarthritis	273	94.79
Trauma	3	1.04
Other	12	4.17

SOME COMMON SEARCHES AND TIPS

Searching for mean age

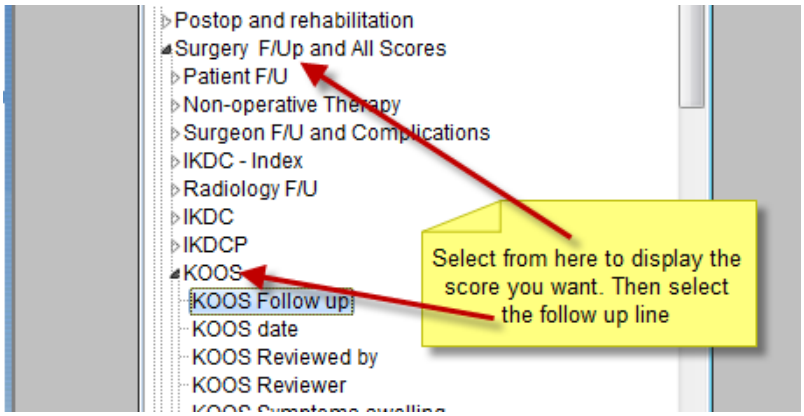
If you are searching for the patients age, don't select the Date of birth as this is meaningless for statistics, what you want is their age at surgery. Once they have their surgery their age stays frozen at that point and they would be 50 with 10 year follow up if their current age is 60. To find this go to the surgery details section, common details and select *Age at Surgery.*



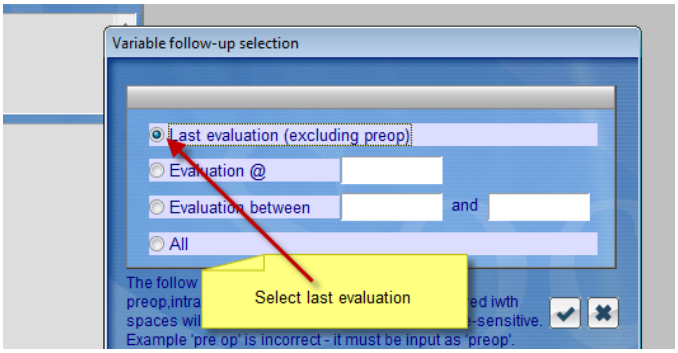
Follow up

This is attached to a score since it calculates the follow up at the time of the score entry. To find your mean and median follow up at the last visit, you first need to know which score you are looking for.

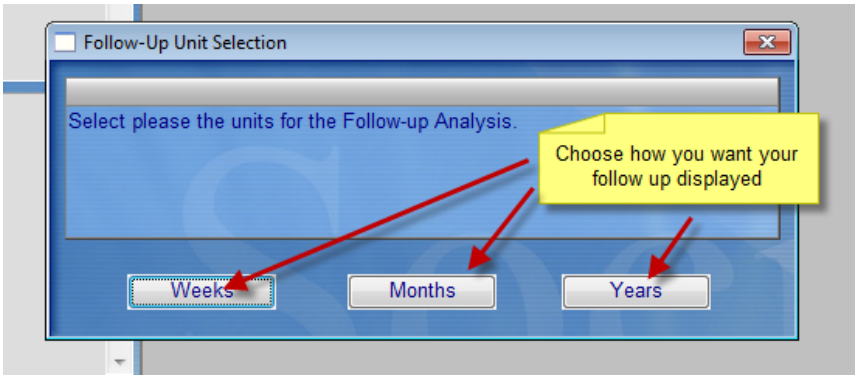
This example shows that we are looking for the mean follow up for the KOOS score. Go the the module you want, select the ***Surgery F/U and All scores***, then select the score you want and then the first line beneath the score name which will be the name of the score and *follow up.*



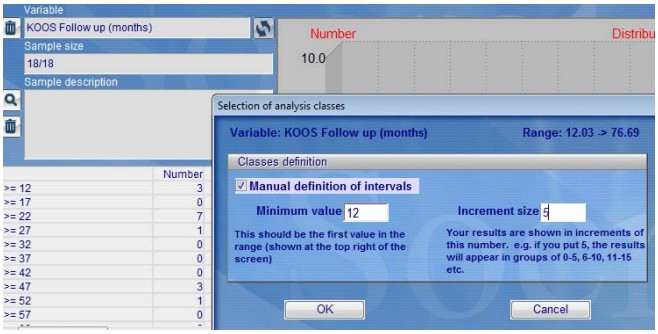
Double click on this and this next window will display. Select last evaluation.



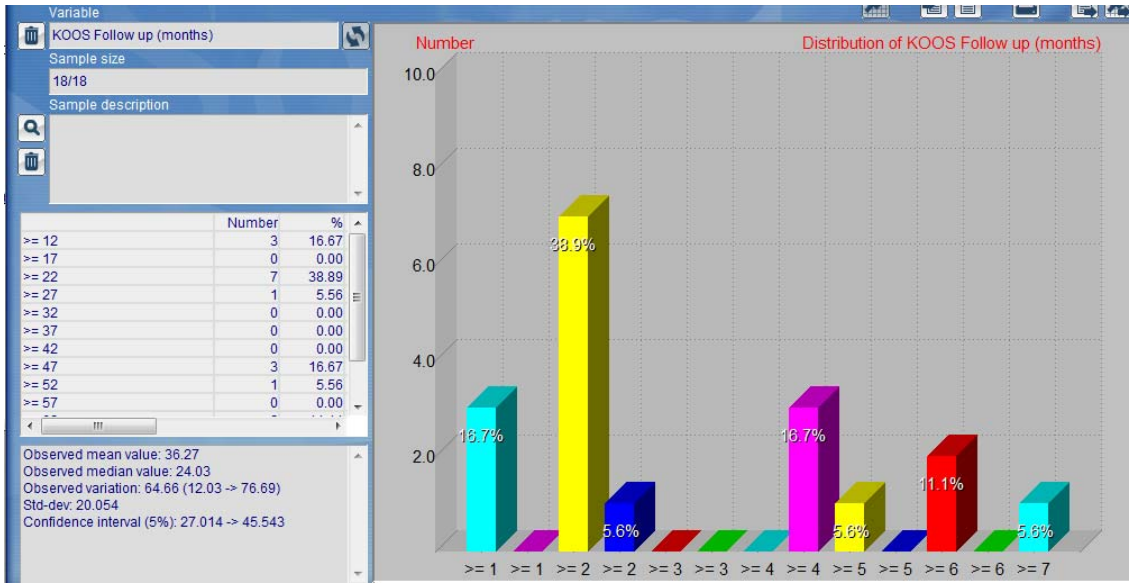
You will then need to choose how you want the follow up displayed – weeks, months or years.



As this is a numeric field you will need to select the range, and increment size.

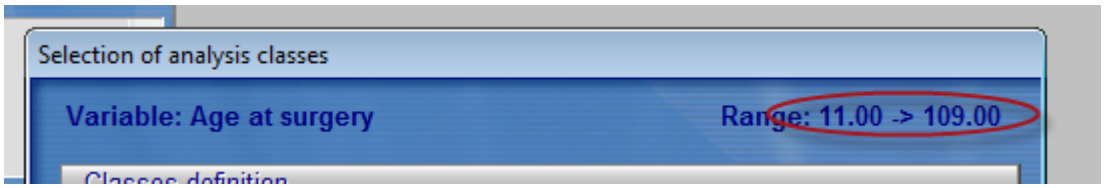


The results in the selected units (months in this example) will then be displayed. In this example the mean follow up is 36 months, the median 24.

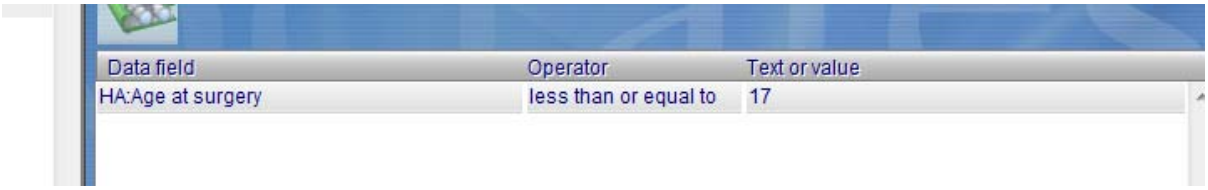


Finding and fixing mistakes

Data entry errors do happen with databases and these are sometimes picked up in the process of running some statistics. For example below we have asked for the age at surgery in the Hip Arthroplasty group and it tells us the range is 11-109.

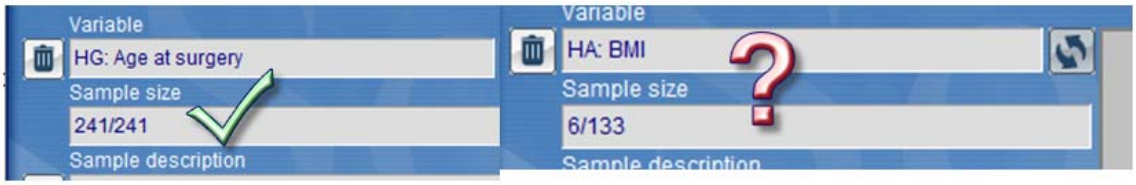


There's possibly a mistake in the DOB or the surgery date (thats how the program calculates the age at surgery) for these entries as its uncommon for an 11 year old to have a hip replacement. And, if you were a sports medicine practice its not common to have 109 year old patients. To find out who these patients are go to the search on the demographic screen and search for all patients whose age at surgery is 11. Or = or < than 17, or more than 100 - whatever range you might want to check on. It will then display the list of those patients by name and you can cross reference their notes for their correct details.



Missing data

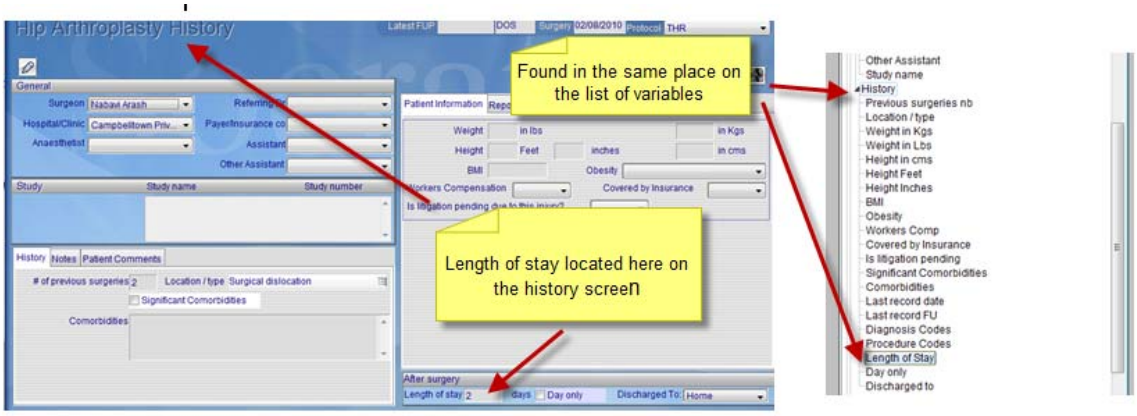
There will always be missing data which may or may not matter depending on what you are studying or think is important. In the examples below the first one had an age for all the patients with records. The second had a BMI for only 6 records. Obviously if you were looking for the effect of BMI and age there would be a problem.



You could go to the search on the demographic screen and use the search parameter *is undefined* to locate the other 127 records which are missing the BMI.

- Make sure you have a clear idea of what you are looking for. It helps to write it down to clarify it before you start looking for data. It's no use asking your secretary to "find your results" for you, you need to be specific about what you want.
- Check the results for obvious data entry errors.
- Don't leave it until the day before your paper is due, there is usually some housekeeping to do and it always takes longer than you think.

- If you don't know your way around the program you will find it hard to find where the variables you are looking for are located. Sometimes it helps to locate the variables you want on the screen first, so you know where to look. For example you are looking for the average length of stay - this is located on the history screen



- The statistics that can be generated here are only descriptive statistics. We will add more as we evolve. You will probably need to export your data to a stats package for more sophisticated statistical analysis correlation and validity tests. Go to the Export chapter for how to do this. We have statisticians who can help you with this also.

HELP IS AVAILABLE!

We have some really clever staff. It is possible to de-identify your database and send it to us and one of our staff can find what you are looking for as well as generate your graphs and even help you write your paper. Ask us for details and a quote.

INDEX