LDMAS (LOH Data Management and Analysis Software) User Guide

Written by : Rifat Hamoudi, Division of Molecular Histopathology, Department of Pathology, University of Cambridge, Cambridge, CB2 2QQ.

Email : rh335@cam.ac.uk

The modules mentioned in this user guide can be downloaded from the following URL :

http://molpath.his.path.cam.ac.uk/bioinformatics/LDMAS.shtml

Introduction

LDMAS package is composed of three modules:

(1) MRES (Medical Report Extractor Software) which parses patient report files, extracts the information of interest and organises it into a structured format, applicable to LDAS.

(2) LDAS (LOH Data Analysis Software) which obtains LOH data from Genotyper and correlates it to clinical data obtained from MRES.

(3) LDMS (LOH Data Management Software) which is used to gather patients' clinico-pathological data and extract significant relationship between the various data sets. LDAS and LDMS work synergistically to manage and analyse LOH data. The MRES source code for automatic parsing of patient reports is written in C++ using C++ Builder 5.0 (Borland Software Corporation, Scotts Valley, CA), LDAS is written in Visual Basic for Application as an Excel 2000 add-in and LDMS is written as Visual Basic for Application modules embedded within Access 2000 as fully functional software. These modules can be run independently and used for applications other than LOH.

Files provided for LDMAS user guide

Filename	**Details**
MRES source code.zip	This contains all the C++ files including the forms and executable file. This software was written using Borland C++ Builder version 5.0
MRES.exe	This is the executable file for MRES software, compiled to run on PC
LDAS.xla	Contains the Visual Basic for Application source code for carrying out the LOH data analysis using Excel platform
LDMS.mdb	Fully functional software that carries out storage and management of data. Also it gives the user flexibility in mining the data
LOH Complete Analysis.xls	Example file of the completed LOH analysis using LDAS
Patient Data after Manual Processing.xls	Example file that was produced by MRES and used as input into LDAS
CIN1_PatientDemo.txt	Example file showing original clinical patient report of 10 patients. This file is used as input by MRES which parses it to produce output similar to "Patient Data after Manual Processing.xls" file.

Example of LDMAS application in identification of LOH markers associated with poor persistence / progression of cervical intraepithelial neoplasia

Cervical intraepithelial neoplasias (CIN) show variable clinical behaviour despite morphological homogeneity within each subgroup. Clinically, it is vital to distinguish CIN lesions with different behaviour and identify those likely to persist and progress despite treatment. We used LDMAS to retrospectively examine the prognostic value of LOH at 12 microsatellite repeat markers including 10 from 3p14, 3p22-21, 6p21 and 11q23 which are frequently deleted in cervical cancer (Giannoudis and Herrington 2001; Lazo 1999), in 164 cases of CIN lesions using archival cytological/histological specimens. LOH was further correlated with high risk HPV infection.

Initially MRES was used to automatically parse 4300 patient records and extract clinico-pathological including age, diagnosis, method of treatment and treatment response during follow up. Out of those 164 cases with follow up of 3 or more years were chosen for the study and their clinico-pathological information was imported into LDAS. Initially, 71 out of the 164 selected cases were examined for LOH using 12 fluorescent microsatellite markers ran on ABI377 DNA Sequencer. LDAS was then used to identify the microsatellite markers for which LOH was significantly associated with disease persistence/progression of CIN using two tailed student t-test. Figure 1b generated using LDAS shows microsatellite markers D3S1300 (3p14.2), D3S1260 (3p22.2), D11S35 (11q22.1) and D11S528 (11q23.3) have the highest LOH in CIN lesions displaying persistence/progression than those displaying disease free during follow up. Stepwise statistical analysis using LDAS showed that concurrent LOH at two of the four microsatellite markers could identify 22-47% CIN displaying disease persistence/progression with 100% specificity. To validate this finding, LOH at these four markers was investigated in a further series of 93 cases. Compatible results were obtained from these additional cases.

The two sets of data were combined and further compared using LDMS. They include 1) comparison of LOH at each of the four microsatellite markers with age, various methods of treatment, different sutypes of HPV infection and between CINs showing disease free or disease persistence/progression, 2) correlation of LOH data with histological grade of CIN, treatment response and various HPV subtypes. Through such complex analysis, we showed that concurrent LOH at two of the four microsatellite markers could identify 22-47% of CINs that showed disease persistence/progression with 100% specificity. Furthermore, LOH at D3S1300 was found to be significantly associated with HPV16 infection. Part of this data analysis is supplied in the LDMAS guide. More detailed analysis of this study is described in (ELhamidi A et al., 2004).

References

ELhamidi, A., Hamoudi, R. A., Kocjan, G., & Du, M. Q. 2004, "Cervical intraepithelial neoplasia: prognosis by combined LOH analysis of multiple loci", *Gynecologic Oncology.*, **94**, 3, 671-679.

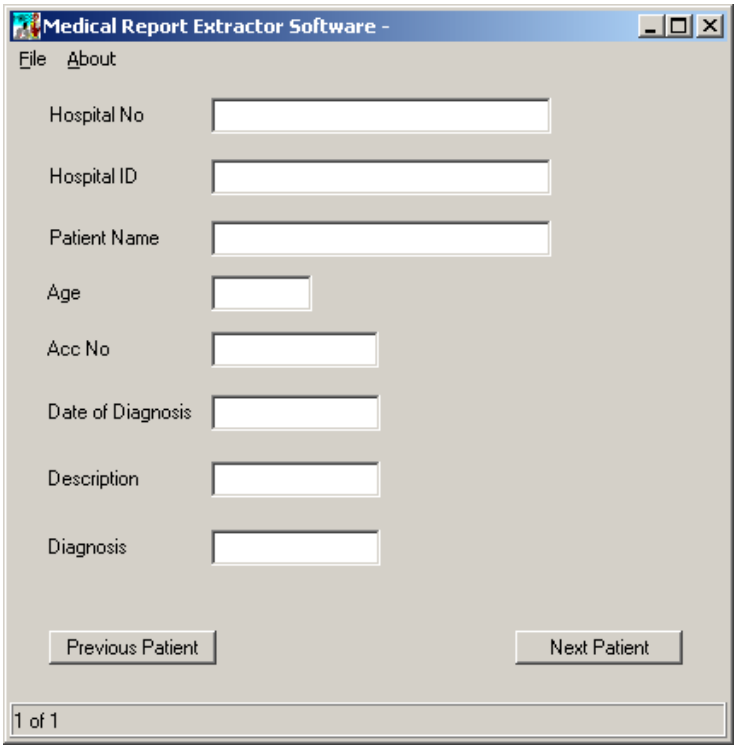
Giannoudis,A. and Herrington,C.S. (2001) Human papillomavirus variants and squamous neoplasia of the cervix. *J.Pathology,* **193,** 295-302.

Lazo,P.A. (1999) The molecular genetics of cervical carcinoma. *British Journal of Cancer,* **80,** 2008-2018.

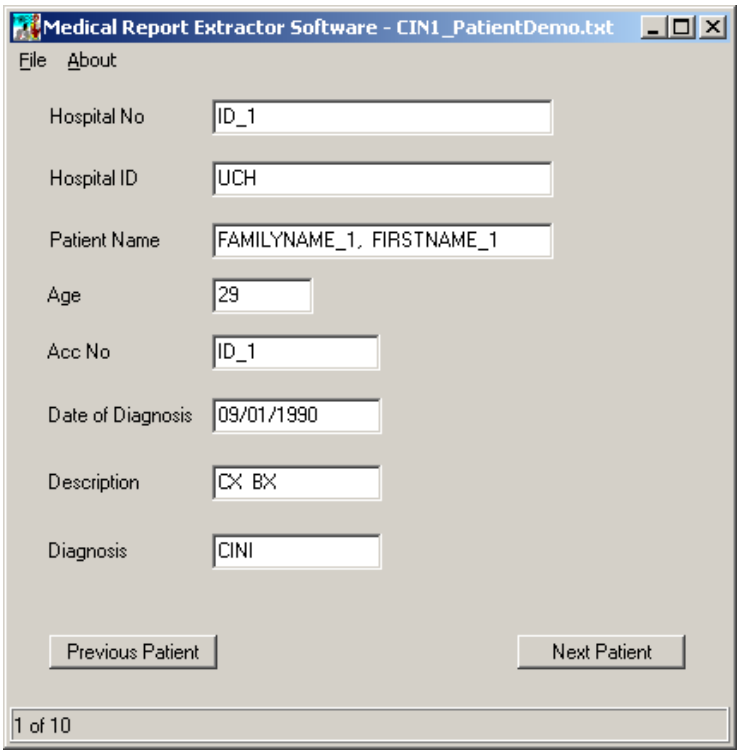
Tamura,S. Nakamori,S. Kuroki,T. Sasaki,Y. Furukawa,H. Ishikawa,O. Imaoka,S. and Nakamura,Y. (1997) Association of cumulative allelic losses with tumor aggressiveness in hepatocellular carcinoma. *Journal of Hepatology,* **27,** 669-676.

(A) Extracting useful clinico-pathological data using MRES (Medical Report Extractor Software)

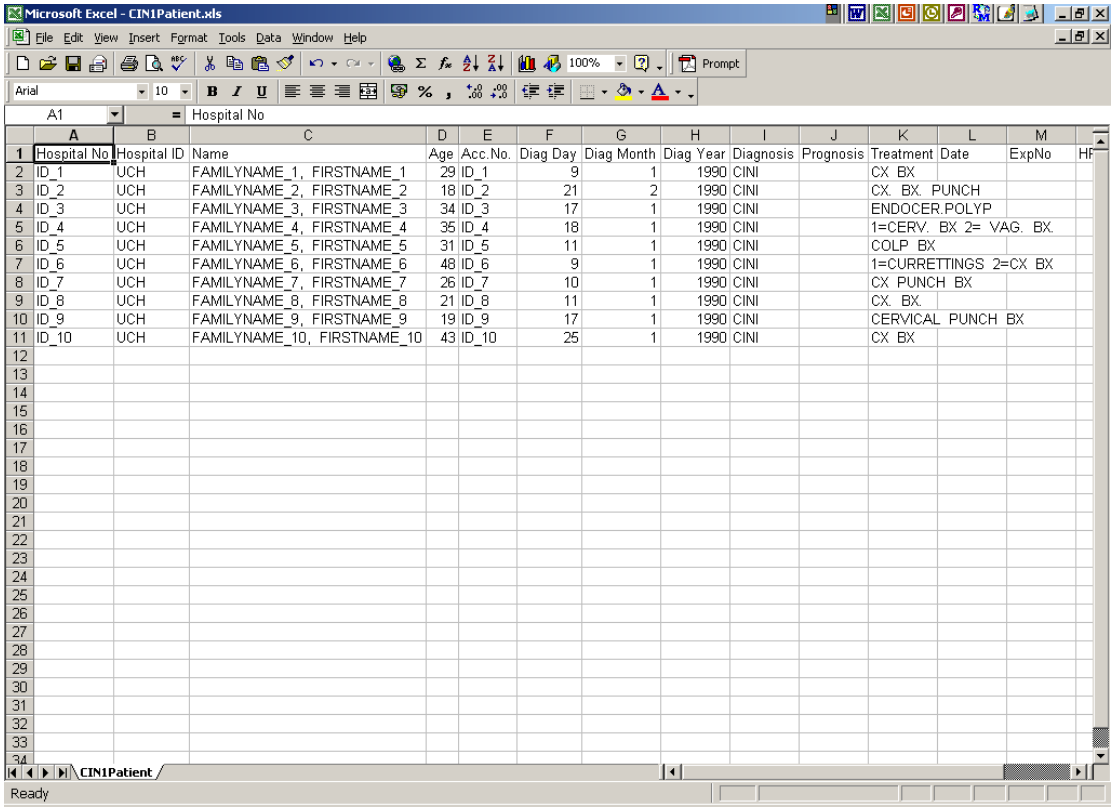
1) Double click on MRES.exe file to get the following GUI:



2) Click on **File** then **Open File** and choose **CIN1_PatientDemo.txt**. You should get the following GUI:



3) To save the file click on **File** then **Save As** and give a filename such as CIN1_PatientDB. The software will save the output in Excel format as default. Therefore you will be able to view the Excel file by double clicking to get the following spreadsheet.



Microsoft Excel - CIN1Patient.xls

File Edit View Insert Format Tools Data Window Help

Arial 10 100% Prompt

A1 = Hospital No

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Hospital No	Hospital ID	Name	Age	Acc.No.	Diag Day	Diag Month	Diag Year	Diagnosis	Prognosis	Treatment	Date	ExpNo
2	ID_1	UCH	FAMILYNAME_1, FIRSTNAME_1	29	ID_1	9	1	1990	CINI		CX BX		
3	ID_2	UCH	FAMILYNAME_2, FIRSTNAME_2	18	ID_2	21	2	1990	CINI		CX. BX. PUNCH		
4	ID_3	UCH	FAMILYNAME_3, FIRSTNAME_3	34	ID_3	17	1	1990	CINI		ENDOCER.POLYP		
5	ID_4	UCH	FAMILYNAME_4, FIRSTNAME_4	35	ID_4	18	1	1990	CINI		1=CERV. BX 2= VAG. BX.		
6	ID_5	UCH	FAMILYNAME_5, FIRSTNAME_5	31	ID_5	11	1	1990	CINI		COLP BX		
7	ID_6	UCH	FAMILYNAME_6, FIRSTNAME_6	48	ID_6	9	1	1990	CINI		1=CURRETTINGS 2=CX BX		
8	ID_7	UCH	FAMILYNAME_7, FIRSTNAME_7	26	ID_7	10	1	1990	CINI		CX PUNCH BX		
9	ID_8	UCH	FAMILYNAME_8, FIRSTNAME_8	21	ID_8	11	1	1990	CINI		CX. BX.		
10	ID_9	UCH	FAMILYNAME_9, FIRSTNAME_9	19	ID_9	17	1	1990	CINI		CERVICAL PUNCH BX		
11	ID_10	UCH	FAMILYNAME_10, FIRSTNAME_10	43	ID_10	25	1	1990	CINI		CX BX		

CIN1Patient

Ready

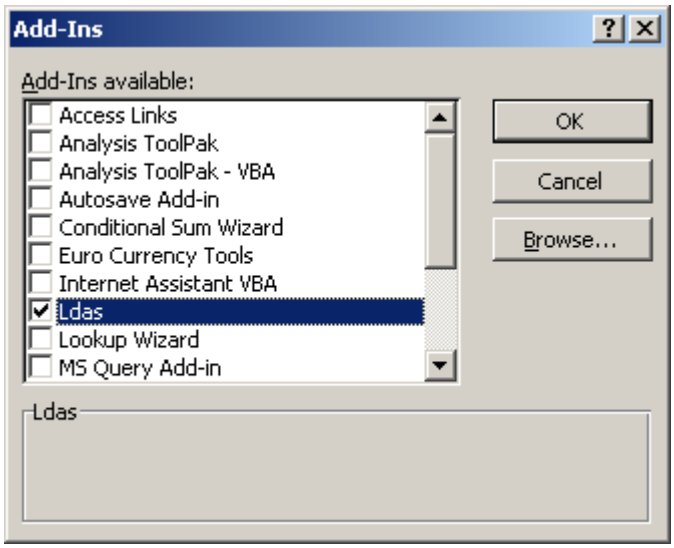
In the LOH study, we produce this sort of file and then manually choose informative cases by examining the follow up time, followed by adding other relevant data such as HPV status. In this example the patient name and hospital ID have been made anonymous and the original clinical report consisted of only 10 patients, however the software can process more than 10,000 patients.

To view the complete annotated file for this study, double click on an Excel file called **Patient Data after Manual Processing.**

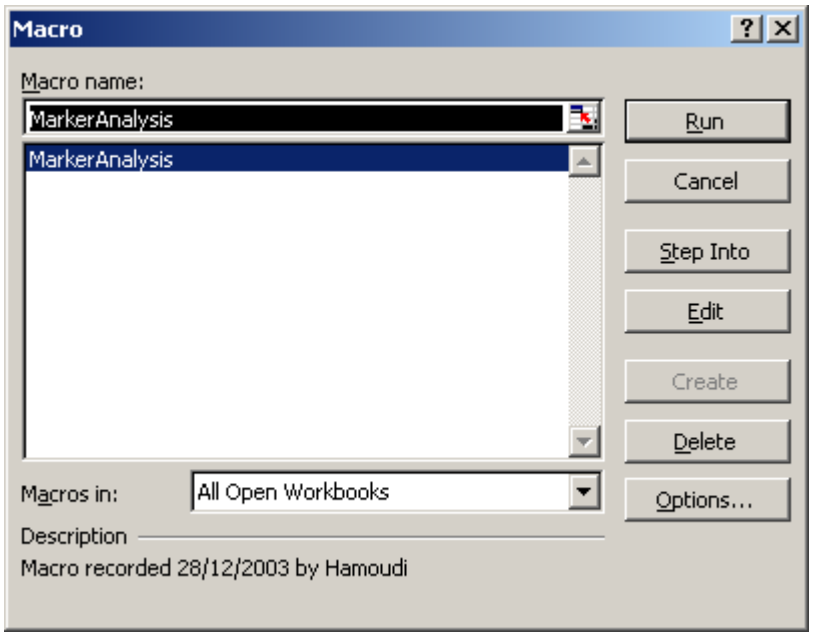
(B) Using LDAS to automate the annotation and analysis of LOH data

1) Double click on **Patient Data after Manual Processing** Excel spreadsheet.

2) Select **Tools** then **Add-Ins**. You should get form as below. Click on **Browse** then select a file called **LDAS.xla**. This should add a module called Ldas to Excel spreadsheet as shown below.

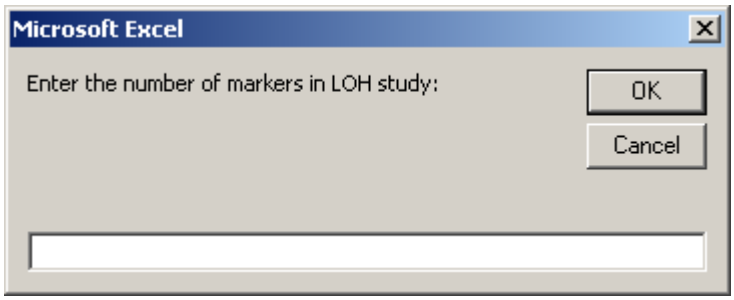


3) From the spreadsheet select **Tools** then **Macros**



Then click on the **Run** option to run a module called **MarkerAnalysis** which is part of the LDAS Add-In.

4) The module should run and it starts by asking you questions about your LOH experiment. The first is to input number of markers (maximum of 4 markers per spreadsheet) as shown below.



Type the marker names

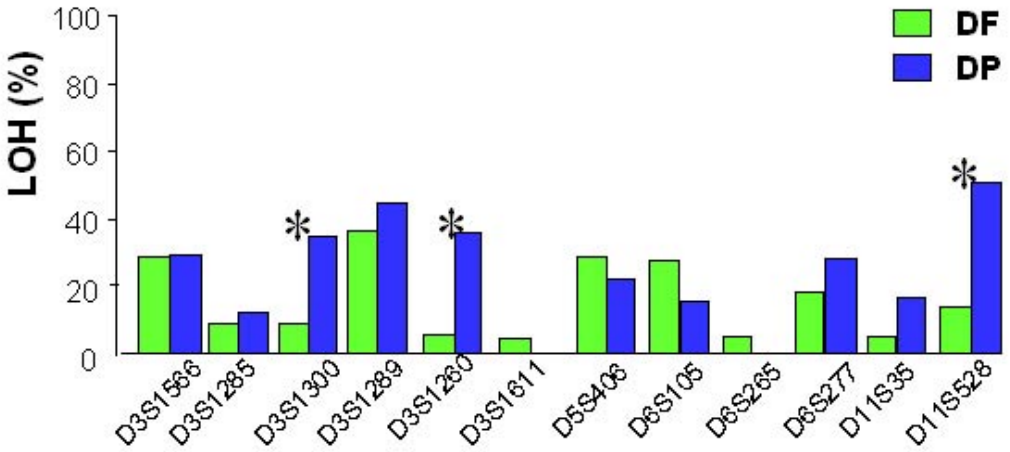
The LDAS can process upto 4 markers for informative and non-informative cases the spreadsheet coordinates are :

```
S1 : Marker 1     yellow (informative)
AX : Marker 1     green (non-informative)

BW : Marker 2     yellow (informative)
DB : Marker 2     green (non-informative)

EA : Marker 3     yellow (informative)
FF : Marker 3     green (non-informative)

GE : Marker 4     yellow (informative)
HJ : Marker 4     green (non-informative)
```

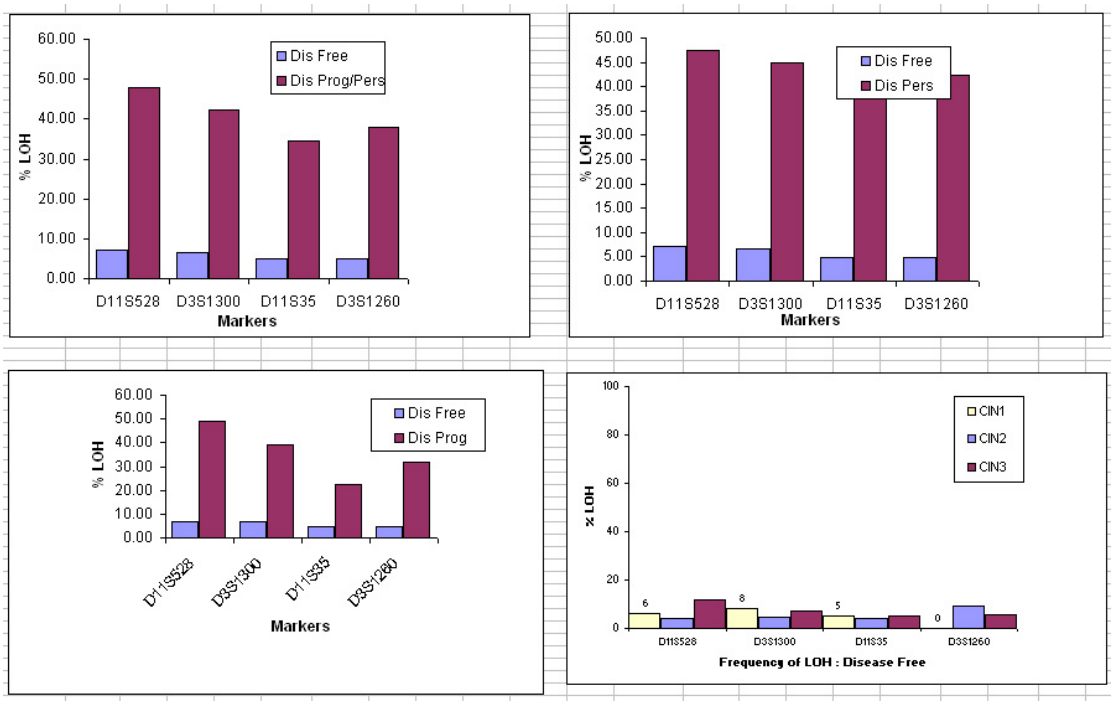
5) Here you can import your data as flat file (plain text format) from either Genotyper (if you used ABI DNA Sequencer to generate LOH data) or GDAS (if you used Affymetrix GeneChip Mapping arrays to generate LOH data).
Once you import your data you can manually edit them within LDAS module in Microsoft Excel if need be.

LDAS will automatically generate the followings :

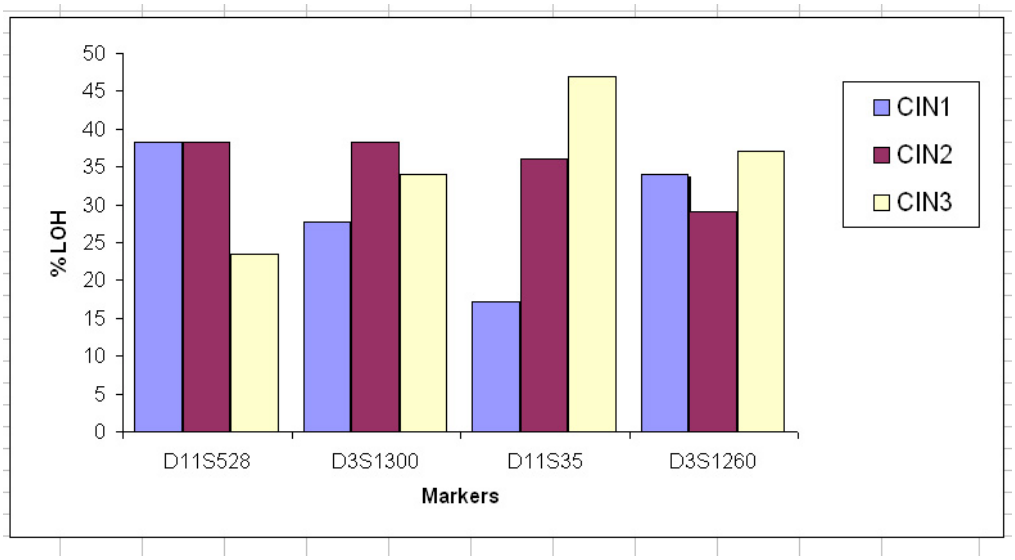
a) Statistics and visualisation of the prognostic markers that can be used to predict the outcome of cervical cancer. For example in the following graph, the most prognostic markers are D3S1300, D3S1260 and D11S528.



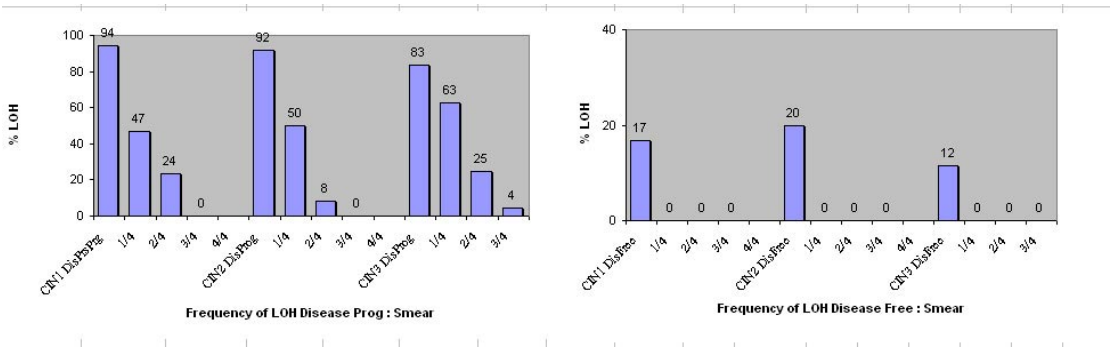
b) The prognostic power of various markers as shown below



c) Stratification of cervical cancer prognostic markers to the various stages of the disease e.g. CIN1, CIN2 and CIN3 as shown below.



d) Sensitivity and specificity of the prognostic markers. This is to test for the degree of false positive using the prognostic markers. The graph below shows that when using 2 markers is sufficient to predict the outcome of cervical cancer with no false positive results. However using 4 markers combined will be too stringent to make accurate prediction.



e) Statistical analysis including correlation of LOH data with clinico-pathological parameters such as treatment, HPV and follow up.

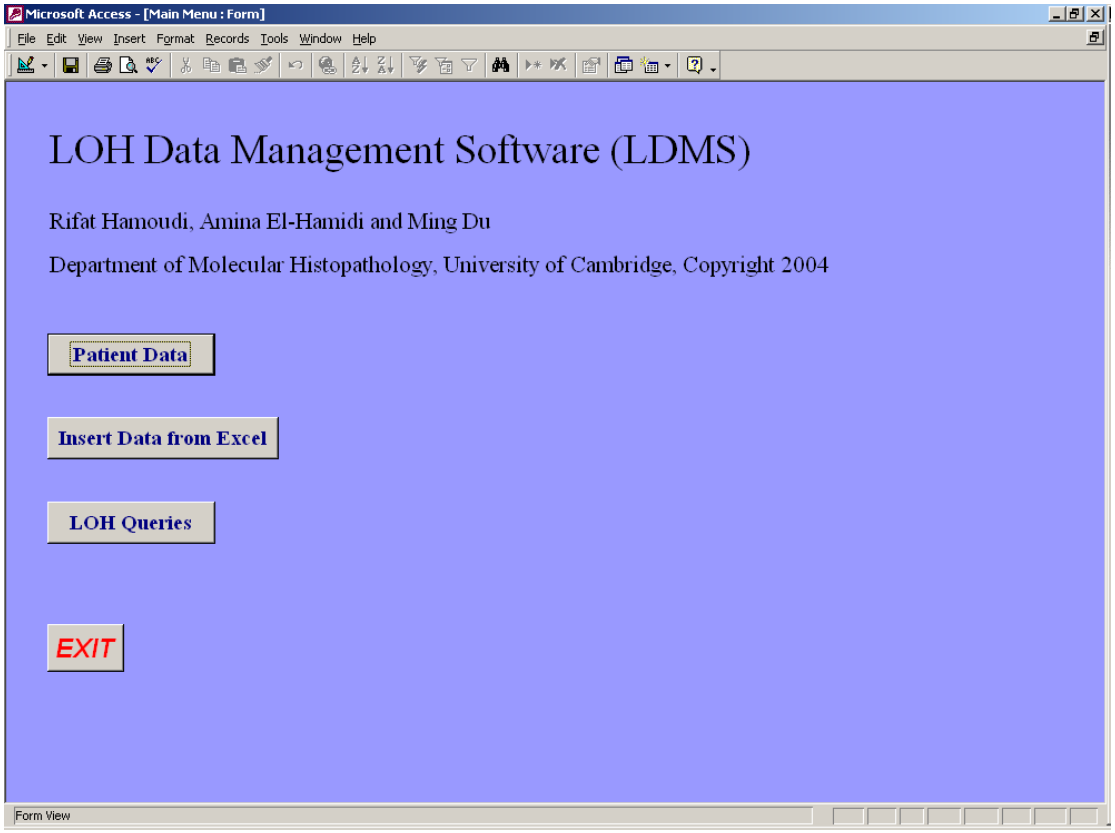
Full results including graphs can be viewed in a file called **LOH Complete Analysis.xls**

The data in this file come from real LOH study that is published in the following paper :

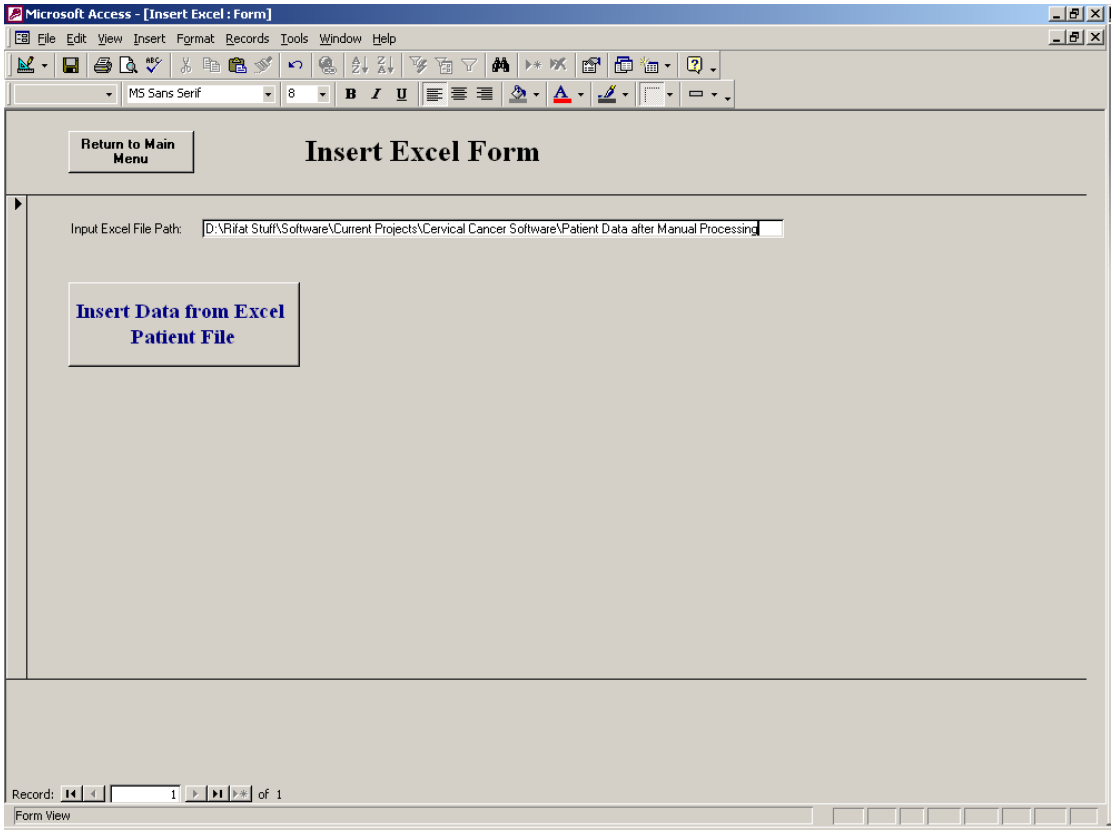
ELhamidi, A., Hamoudi, R. A., Kocjan, G., & Du, M. Q. 2004, "Cervical intraepithelial neoplasia: prognosis by combined LOH analysis of multiple loci", *Gynecologic Oncology.*, **94**, 3, 671-679.

(C) Using LDMS to store, manage and mine the clinico-pathological data

1) Double click on LDMS module. You should get the following :



2) To enter data click on **Insert Data from Excel**. You should get the following :

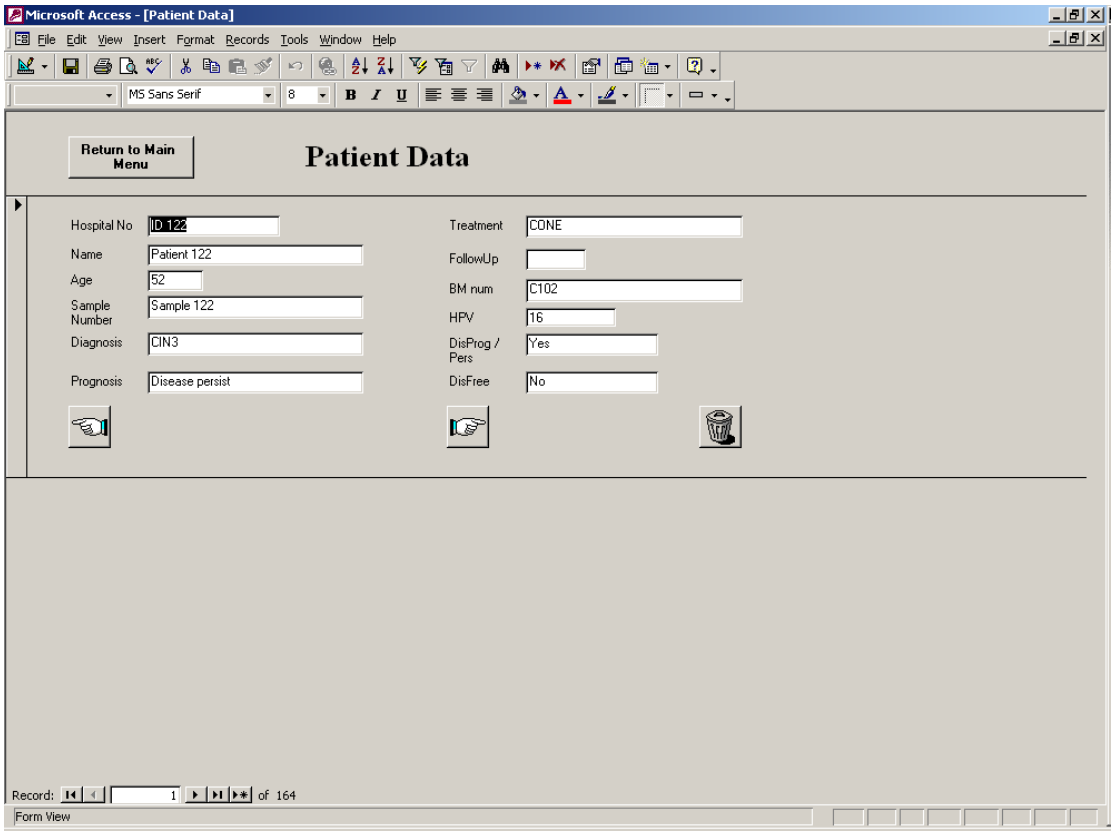


Type the full path of the Excel patient database and add the filename.

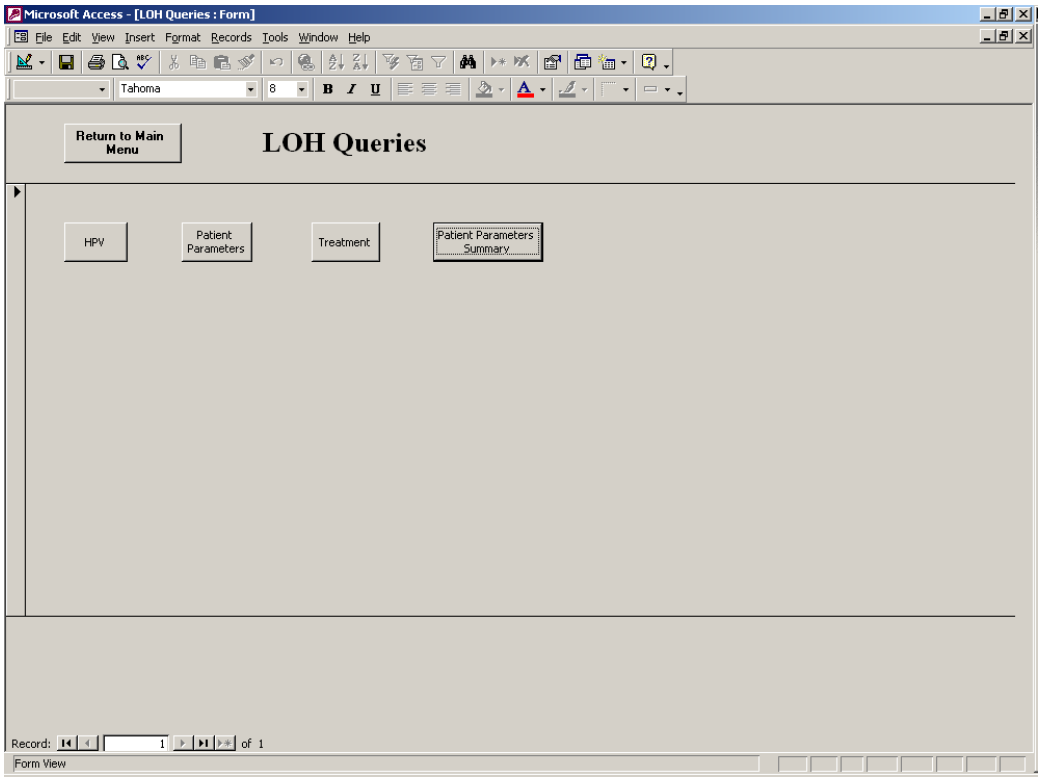
The filename in this case is : **Patient Data after Manual Processing**

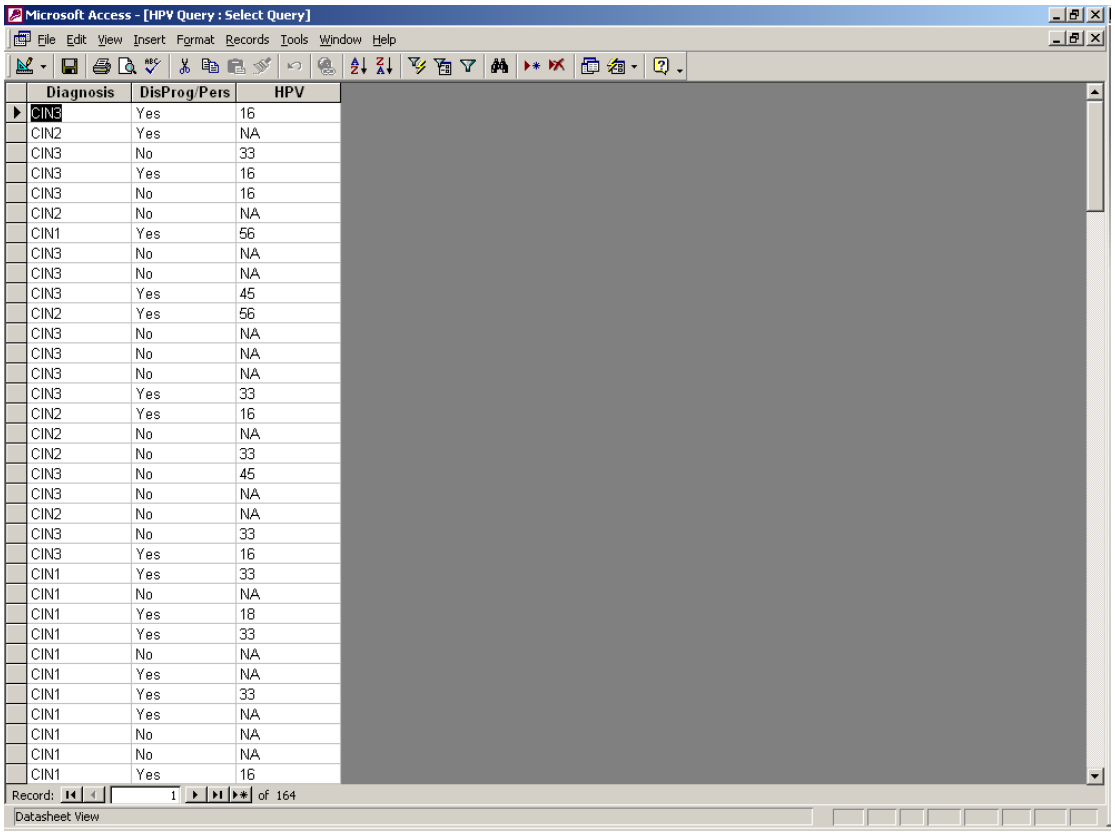
When successful you should get the following window message

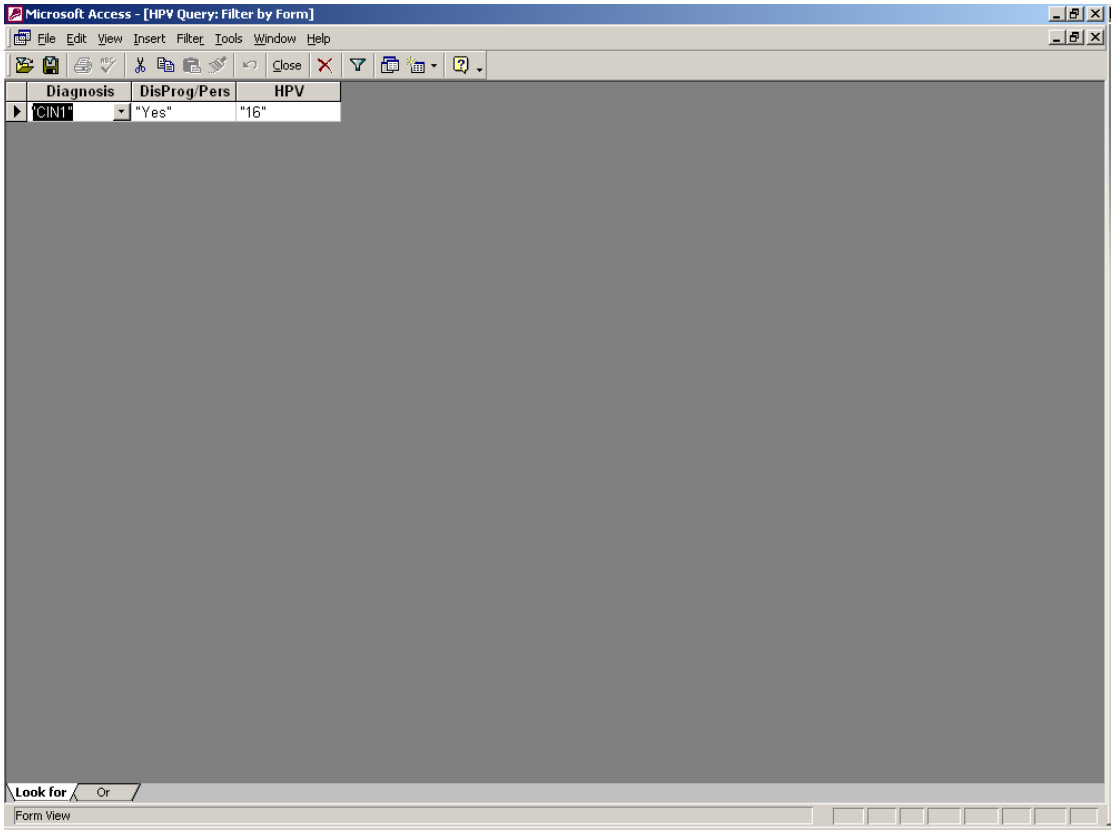
3) Check that the patient data are entered by clicking on **Patient Data** and moving to next record

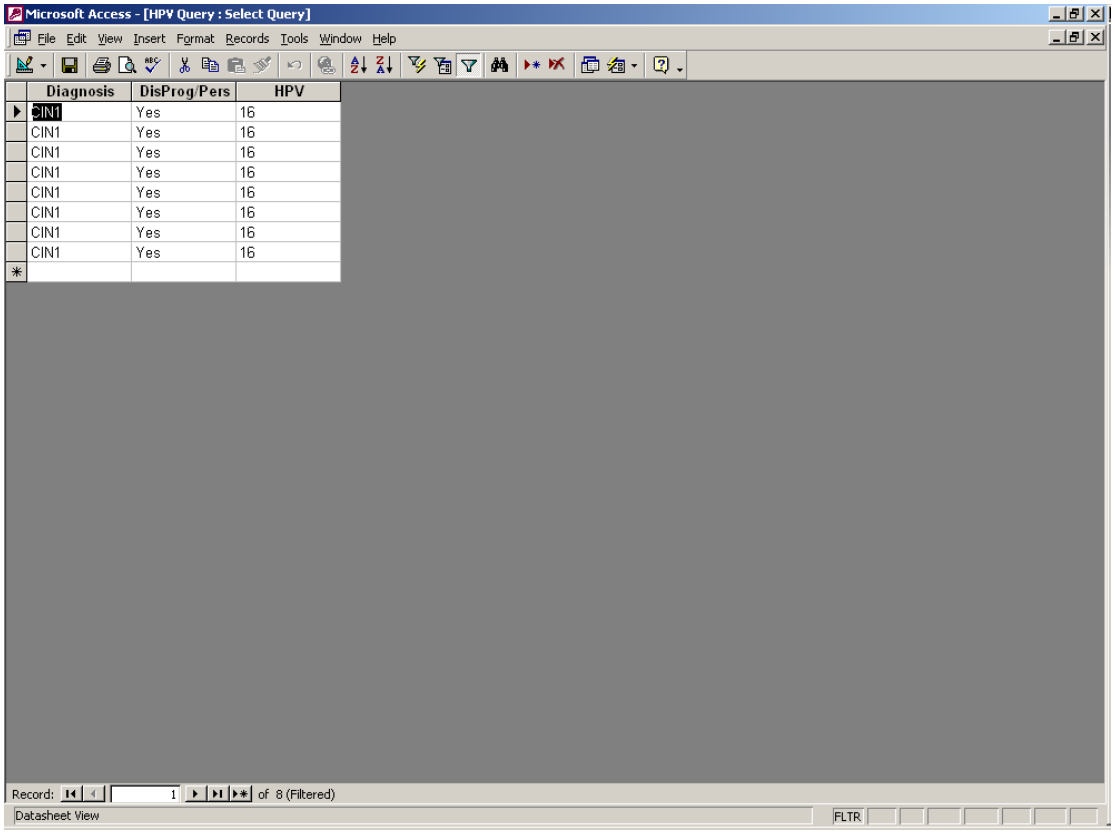


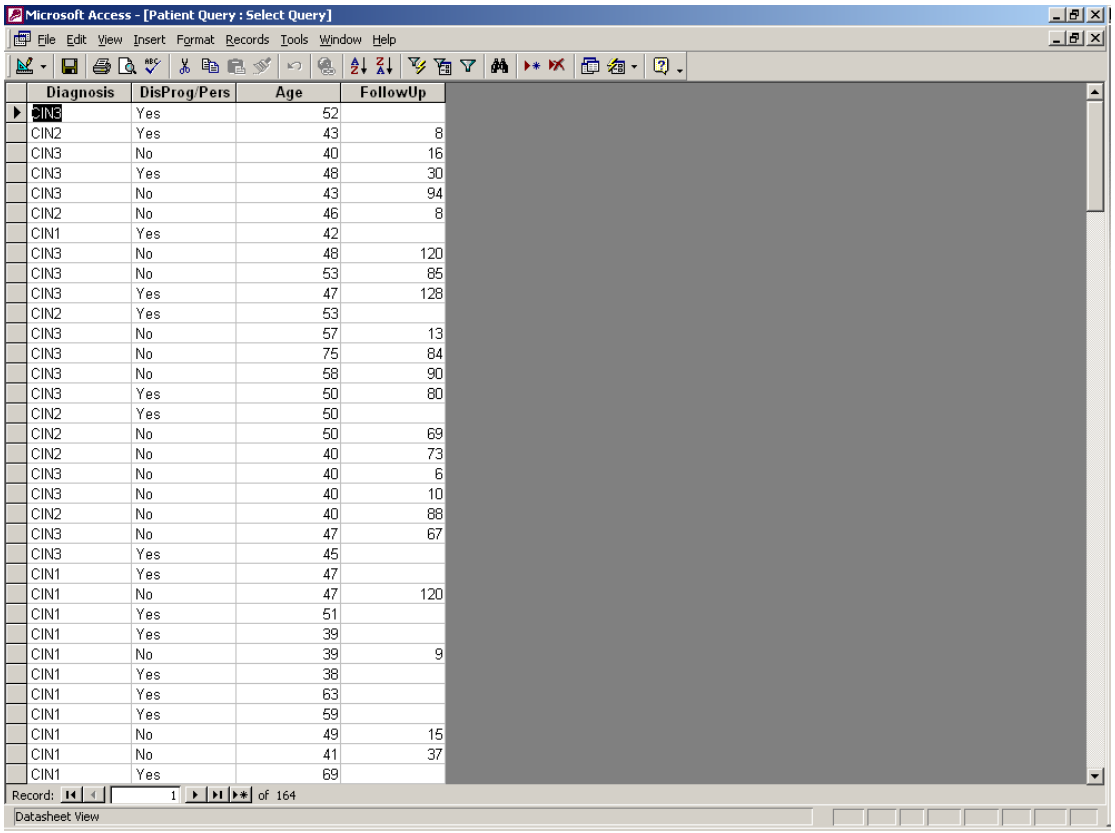
4) Once entered click on **LOH Queries** to carry out data mining.

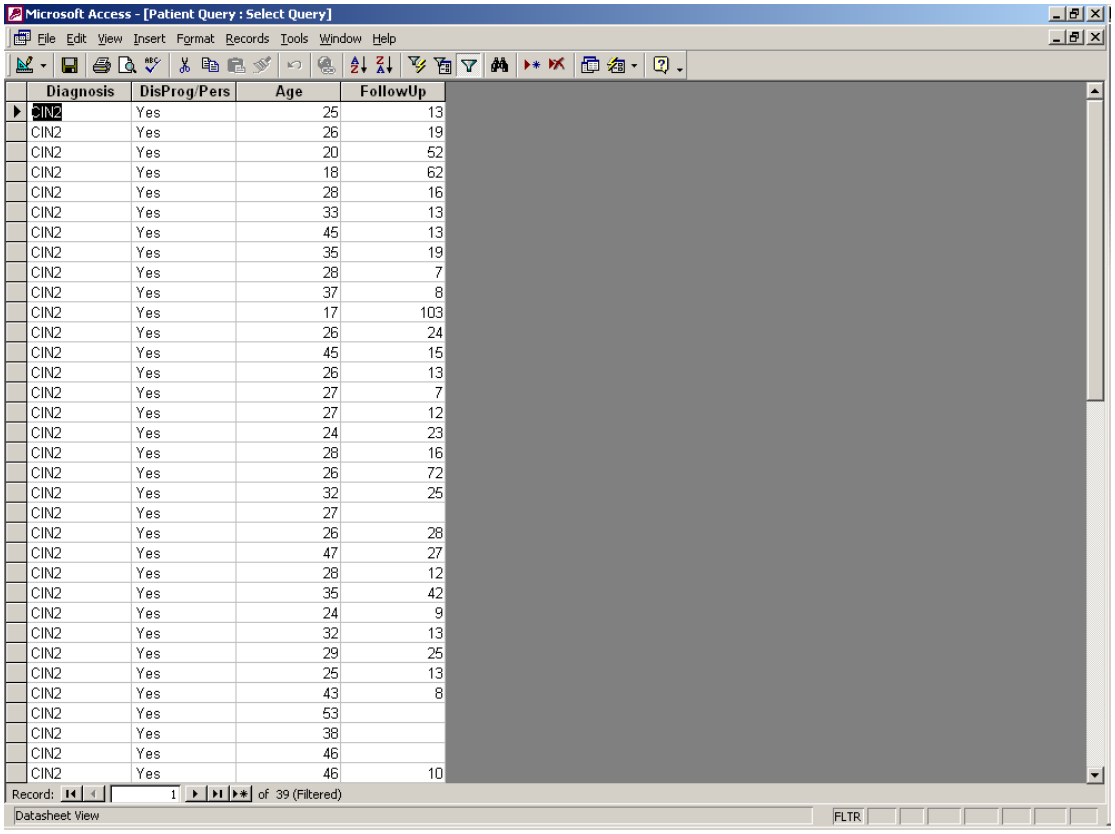


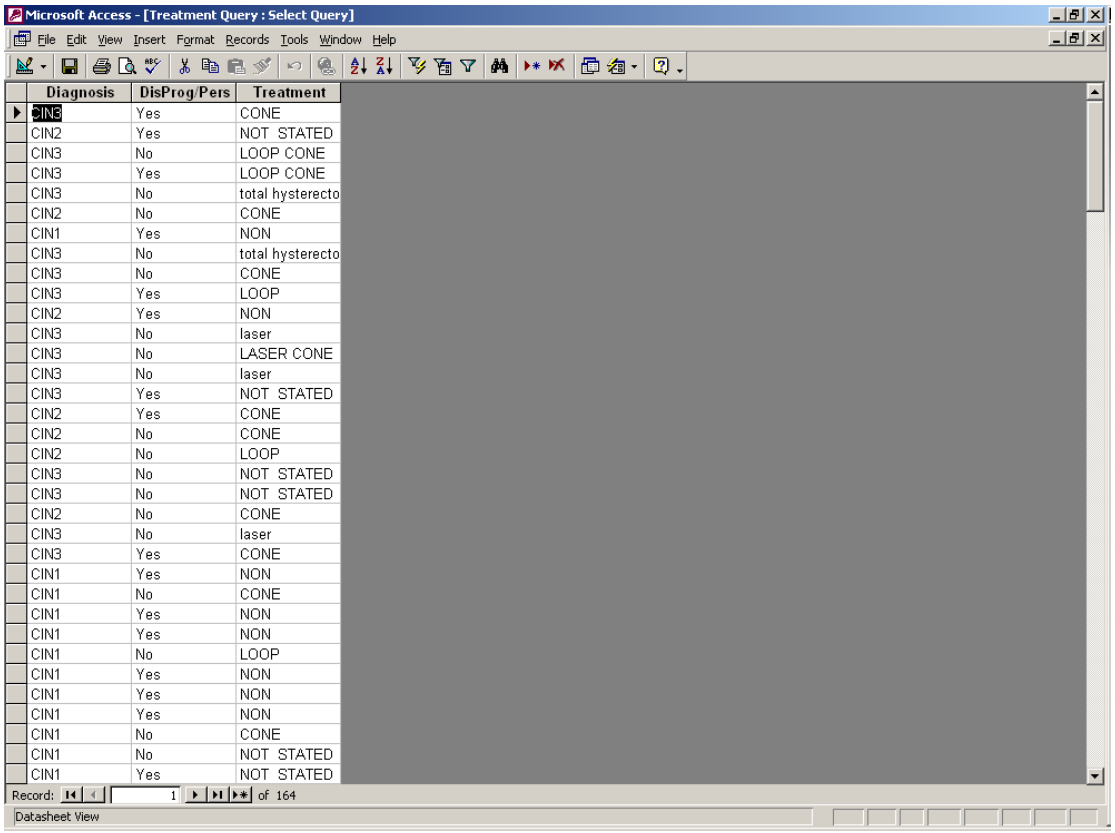

Microsoft Access - [HPV Query : Select Query]
File Edit View Insert Format Records Tools Window Help
Diagnosis DisProg/Pers HPV
CIN3 Yes 16
CIN2 Yes NA
CIN3 No 33
CIN3 Yes 16
CIN3 No 16
CIN2 No NA
CIN1 Yes 56
CIN3 No NA
CIN3 No NA
CIN3 Yes 45
CIN2 Yes 56
CIN3 No NA
CIN3 No NA
CIN3 No NA
CIN3 Yes 33
CIN2 Yes 16
CIN2 No NA
CIN2 No 33
CIN3 No 45
CIN3 No NA
CIN2 No NA
CIN3 No 33
CIN3 Yes 16
CIN1 Yes 33
CIN1 No NA
CIN1 Yes 18
CIN1 Yes 33
CIN1 No NA
CIN1 Yes NA
CIN1 Yes 33
CIN1 Yes NA
CIN1 No NA
CIN1 No NA
CIN1 Yes 16
Record: 1 of 164
Datasheet View

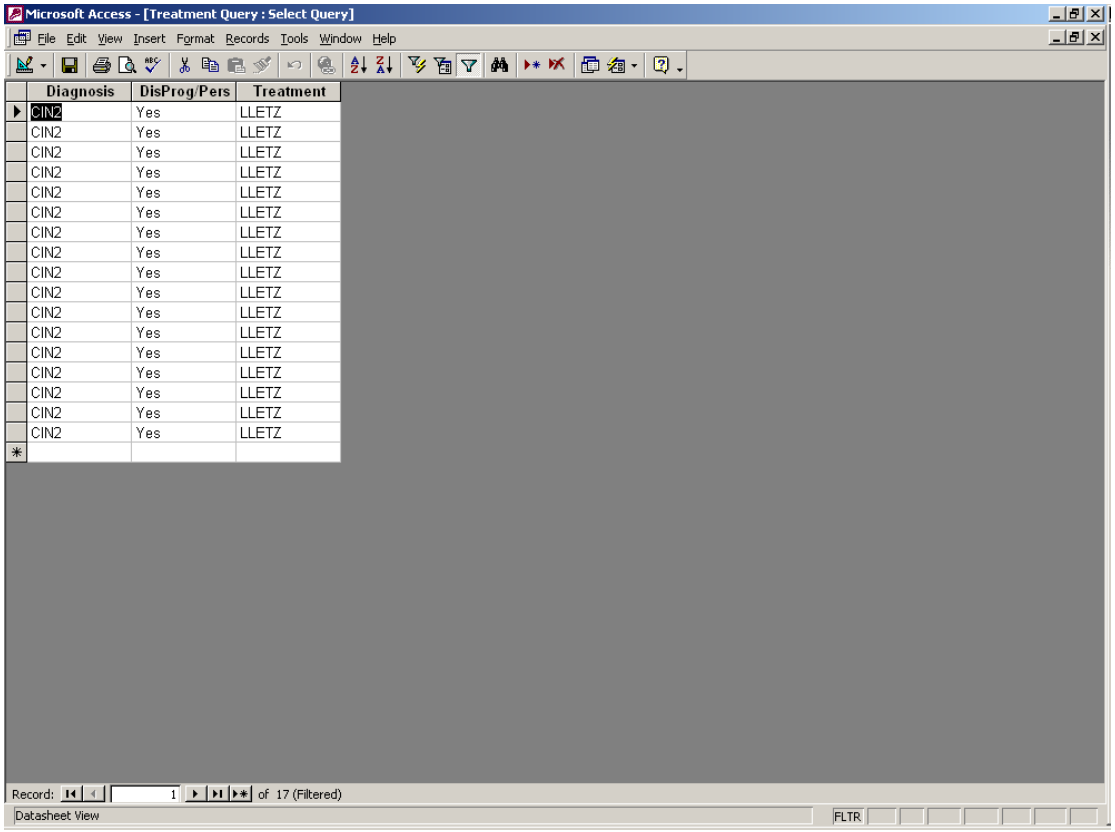

Microsoft Access - [HPV Query: Filter by Form]
File Edit View Insert Filter Tools Window Help
Close
Diagnosis DisProg/Pers HPV
"CIN1" "Yes" "16"
Look for Or
Form View


Microsoft Access - [HPV Query : Select Query]
File Edit View Insert Format Records Tools Window Help
Diagnosis | DisProg/Pers | HPV
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
CIN1 | Yes | 16
Record: 1 of 8 (Filtered)
Datasheet View
FLTR

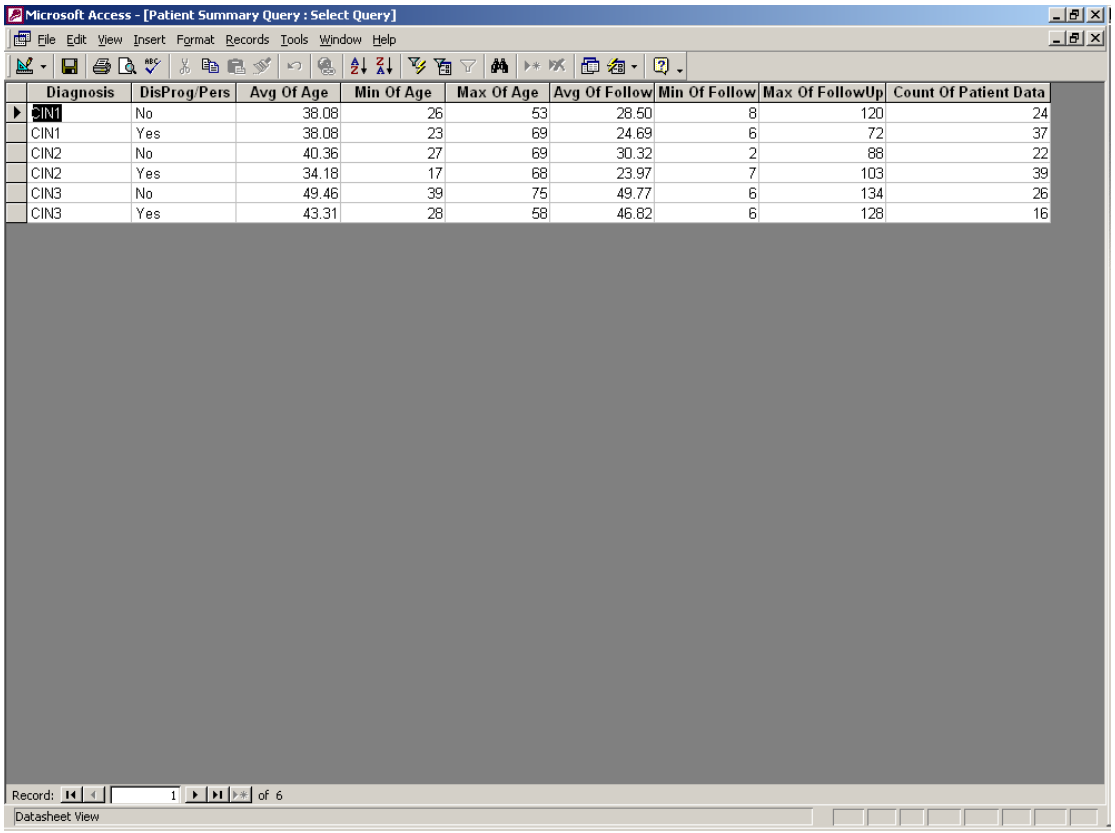

Microsoft Access - [Patient Query : Select Query]
File Edit View Insert Format Records Tools Window Help
Diagnosis | DisProg/Pers | Age | FollowUp
CIN3 | Yes | 52
CIN2 | Yes | 43 | 8
CIN3 | No | 40 | 16
CIN3 | Yes | 48 | 30
CIN3 | No | 43 | 94
CIN2 | No | 46 | 8
CIN1 | Yes | 42
CIN3 | No | 48 | 120
CIN3 | No | 53 | 85
CIN3 | Yes | 47 | 128
CIN2 | Yes | 53
CIN3 | No | 57 | 13
CIN3 | No | 75 | 84
CIN3 | No | 58 | 90
CIN3 | Yes | 50 | 80
CIN2 | Yes | 50
CIN2 | No | 50 | 69
CIN2 | No | 40 | 73
CIN3 | No | 40 | 6
CIN3 | No | 40 | 10
CIN2 | No | 40 | 88
CIN3 | No | 47 | 67
CIN3 | Yes | 45
CIN1 | Yes | 47
CIN1 | No | 47 | 120
CIN1 | Yes | 51
CIN1 | Yes | 39
CIN1 | No | 39 | 9
CIN1 | Yes | 38
CIN1 | Yes | 63
CIN1 | Yes | 59
CIN1 | No | 49 | 15
CIN1 | No | 41 | 37
CIN1 | Yes | 69
Record: 1 of 164
Datasheet View


Microsoft Access - [Patient Query : Select Query]
File Edit View Insert Format Records Tools Window Help
Diagnosis | DisProg/Pers | Age | FollowUp
CIN2 | Yes | 25 | 13
CIN2 | Yes | 26 | 19
CIN2 | Yes | 20 | 52
CIN2 | Yes | 18 | 62
CIN2 | Yes | 28 | 16
CIN2 | Yes | 33 | 13
CIN2 | Yes | 45 | 13
CIN2 | Yes | 35 | 19
CIN2 | Yes | 28 | 7
CIN2 | Yes | 37 | 8
CIN2 | Yes | 17 | 103
CIN2 | Yes | 26 | 24
CIN2 | Yes | 45 | 15
CIN2 | Yes | 26 | 13
CIN2 | Yes | 27 | 7
CIN2 | Yes | 27 | 12
CIN2 | Yes | 24 | 23
CIN2 | Yes | 28 | 16
CIN2 | Yes | 26 | 72
CIN2 | Yes | 32 | 25
CIN2 | Yes | 27
CIN2 | Yes | 26 | 28
CIN2 | Yes | 47 | 27
CIN2 | Yes | 28 | 12
CIN2 | Yes | 35 | 42
CIN2 | Yes | 24 | 9
CIN2 | Yes | 32 | 13
CIN2 | Yes | 29 | 25
CIN2 | Yes | 25 | 13
CIN2 | Yes | 43 | 8
CIN2 | Yes | 53
CIN2 | Yes | 38
CIN2 | Yes | 46
CIN2 | Yes | 46 | 10
Record: 1 of 39 (Filtered)
Datasheet View
FLTR


Microsoft Access - [Treatment Query : Select Query]
File Edit View Insert Format Records Tools Window Help
Diagnosis | DisProg/Pers | Treatment
CIN3 | Yes | CONE
CIN2 | Yes | NOT STATED
CIN3 | No | LOOP CONE
CIN3 | Yes | LOOP CONE
CIN3 | No | total hysterecto
CIN2 | No | CONE
CIN1 | Yes | NON
CIN3 | No | total hysterecto
CIN3 | No | CONE
CIN3 | Yes | LOOP
CIN2 | Yes | NON
CIN3 | No | laser
CIN3 | No | LASER CONE
CIN3 | No | laser
CIN3 | Yes | NOT STATED
CIN2 | Yes | CONE
CIN2 | No | CONE
CIN2 | No | LOOP
CIN3 | No | NOT STATED
CIN3 | No | NOT STATED
CIN2 | No | CONE
CIN3 | No | laser
CIN3 | Yes | CONE
CIN1 | Yes | NON
CIN1 | No | CONE
CIN1 | Yes | NON
CIN1 | Yes | NON
CIN1 | No | LOOP
CIN1 | Yes | NON
CIN1 | Yes | NON
CIN1 | Yes | NON
CIN1 | No | CONE
CIN1 | No | NOT STATED
CIN1 | Yes | NOT STATED
Record: 1 of 164
Datasheet View



Diagnosis	DisProg/Pers	Treatment
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ
CIN2	Yes	LLETZ



Diagnosis	DisProg/Pers	Avg Of Age	Min Of Age	Max Of Age	Avg Of Follow	Min Of Follow	Max Of FollowUp	Count Of Patient Data
CIN1	No	38.08	26	53	28.50	8	120	24
CIN1	Yes	38.08	23	69	24.69	6	72	37
CIN2	No	40.36	27	69	30.32	2	88	22
CIN2	Yes	34.18	17	68	23.97	7	103	39
CIN3	No	49.46	39	75	49.77	6	134	26
CIN3	Yes	43.31	28	58	46.82	6	128	16

This example shows how to mine the data and extract useful clinical information from it such as HPV status, treatment and follow up details.