Case-Control study

Problem

- 200 students appeared for Exam
- Only 90 passed, 110 failed
- Problem – Why many students failed?

What is the reason?

- Analysis – Possible reasons for failure
  1. Question paper difficult - No
  2. Lecture taken – Yes
  3. Attended lecture – All students did not attend.
- How do we know not attending the lecture is the correct reason?

Case Control study

Reason	Failed in exam (Problem +)	Pass (Problem -)	
Reason + (absent from lecture)	a	b	a+b
Reason '–' (present for lecture)	c	d	c+d
	a+c	b+d	

	Failed in exam (problem+)	Pass (Problem-)	
Reason '+' (**absent** from lecture)			a+b
Reason '–' (present for lecture)			c+d
	110 (a+c)	90 (b+d)	200

	Failed in exam (problem+)	Pass (Problem-)	
Reason '+' (**absent** from lecture)	100 (a)	10 (b)	a+b
Reason '–' (present for lecture)	10 (c)	80 (d)	c+d
	110 (a+c)	90 (b+d)	200

- ***Definition***

The Odds Ratio is a measure of ***association*** which compares the odds of disease of those exposed **(cases)** to the odds of disease those unexposed **(control).**

- ***Formulae***

OR = (odds of disease in exposed) / (odds of disease in the non-exposed)

odds of exposure in cases= No of cases with exposure/ No of cases without exposure = **a/c**

odds of exposure in control= No of controls with exposure/ No of controls without exposure = **b/d**

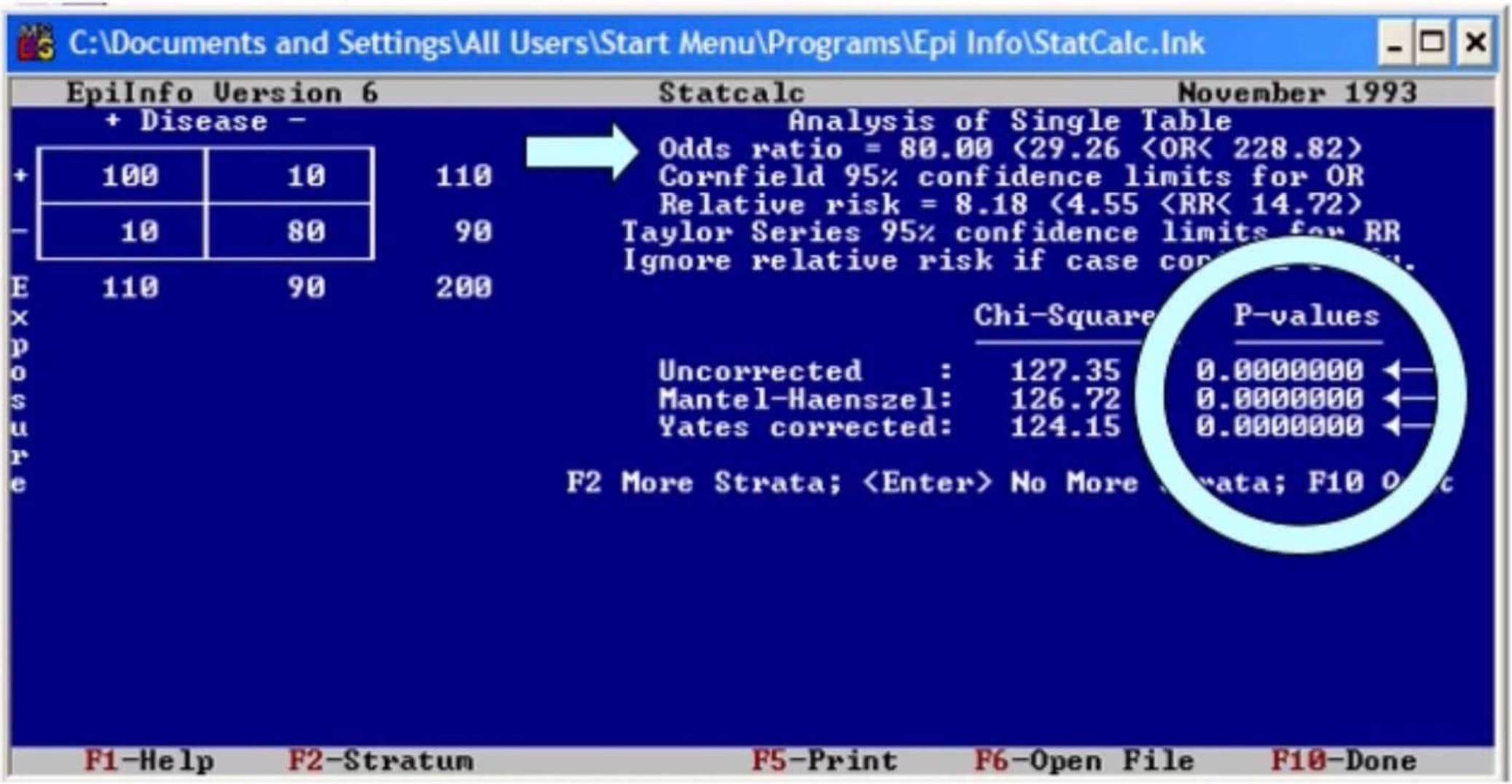
$$\frac{\text{odds of exposure in cases}}{\text{odds of exposure in controls}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

$$\text{OR} = \frac{ad}{bc}$$

Analysis

- Odds ratio = axd/bxc
- 100x80/10x10

  = 80
- **Students not attending lecture has 80 times more chance of failing in the exam than those who attend the lecture.**



This result is not by chance because P-value is < (less than) 0.05

Phocomelia

Phocomelia

- In 1950 many children in Europe were born with Phocomelia
- Doctors were worried. Why is this happening?
- They asked mothers of these children
  - Was there any problem during pregnancy? No
  - Did they suffer from any disease? No
  - Did they take any medicine? Yes – Thalidomide for morning sickness

Thalidomide tragedy

- They did a Case-Control study to find whether it is the reason
- Case-Control study proved that Thalidomide was the cause
- Thalidomide was banned

Analytical epidemiology

- We test whether there is an association between a disease and the suspected factor.
- We also measure the strength of association.

Case Control study

- Sometimes called 'retrospective study'
- Is the first step to test hypothesis
- Both cause and outcome (disease) have occurred before doing the study

Case Control study

- There is a control or comparison group to test the hypothesis
- This is the most important feature of Analytical epidemiology

Design of a Case Control study

	Cases (Disease+)	Control (Disease-)	
Risk factor present(+)	a	b	
Risk factor absent (-)	c	d	
	a+c	b+d	

Design of Case Control study

	Cases (Disease +)	Control (Disease -)	
Risk factor +	a	b	
Risk factor -	c	d	
	a+c	b+d	

Steps in a Case Control study

- Selection of cases and controls
- Matching
- Measurement of exposure to risk factor
- Analysis and interpretation

Selection of Cases

- Cases selected should have the correct diagnosis.
- Only cases with the confirmed diagnosis should be included

Selection of Controls

- Controls must be **FREE** from the disease under study.
- If there are sub-clinical cases, do laboratory test to make sure that the person has no disease

Sources of controls

- Hospitals (patients having other disease)
- Neighborhood controls
- General population

- How many controls will you take for a case?
  - In large studies generally 1
  - In small studies (below 50) up to 4

Matching

- Matching is a process by which we select controls in a such a way that they are similar to cases in important variables
- Age, Sex, Occupation etc.
- By matching we can neutralize any confounding factor.

Matching - examples

- For studying Lung cancer the Controls should be males and **not** females
- For studying Lung cancer the Controls should be adult males and **not** small boys
- For studying Breast cancer the controls should be females and **not** males!
- For studying Breast cancer the controls should be adult females and **not** small girls

Measurement of exposure to cause

- There must be a clear Definition for the risk factor.
- That should be same for Cases and Controls
- E.g. Smoking- number of cigarettes, duration of smoking, type of cigarette etc.

Analysis

- Calculate exposure rates among cases and controls
- Calculate the disease risk associated with exposure (Odds ratio)

Analysis

	Cases (Lung cancer +)	Controls (No lung cancer)	
Smoking +	33(a)	55(b)	
Smoking -	2(c)	27(d)	
	35(a+c)	82(b+d)	

Exposure rate to smoking

Cases = a/a+c 33/35 = 94.2%

Controls = b/b+d 55/82 = 67%

Estimation of risk

- Those who are having lung cancer are smoking more(94.2%)
- However it does **not** mean that 94.2% of all smokers will develop lung cancer.
- We estimate risk to develop lung cancer in smokers by calculating 'Odds ratio'

Odds ratio

Odds ratio = ad/bc

33x27/55x2 = 8.1

Those who smoke have 8.1 times the risk of developing Lung cancer than those who do not smoke

- If the odds ratio is 1 means **no** risk or exposure doesn't affect odds of the disease.
- OR > 1 exposure associated with higher odds of disease.
- OR< 1 exposure associated with lower odds of disease.

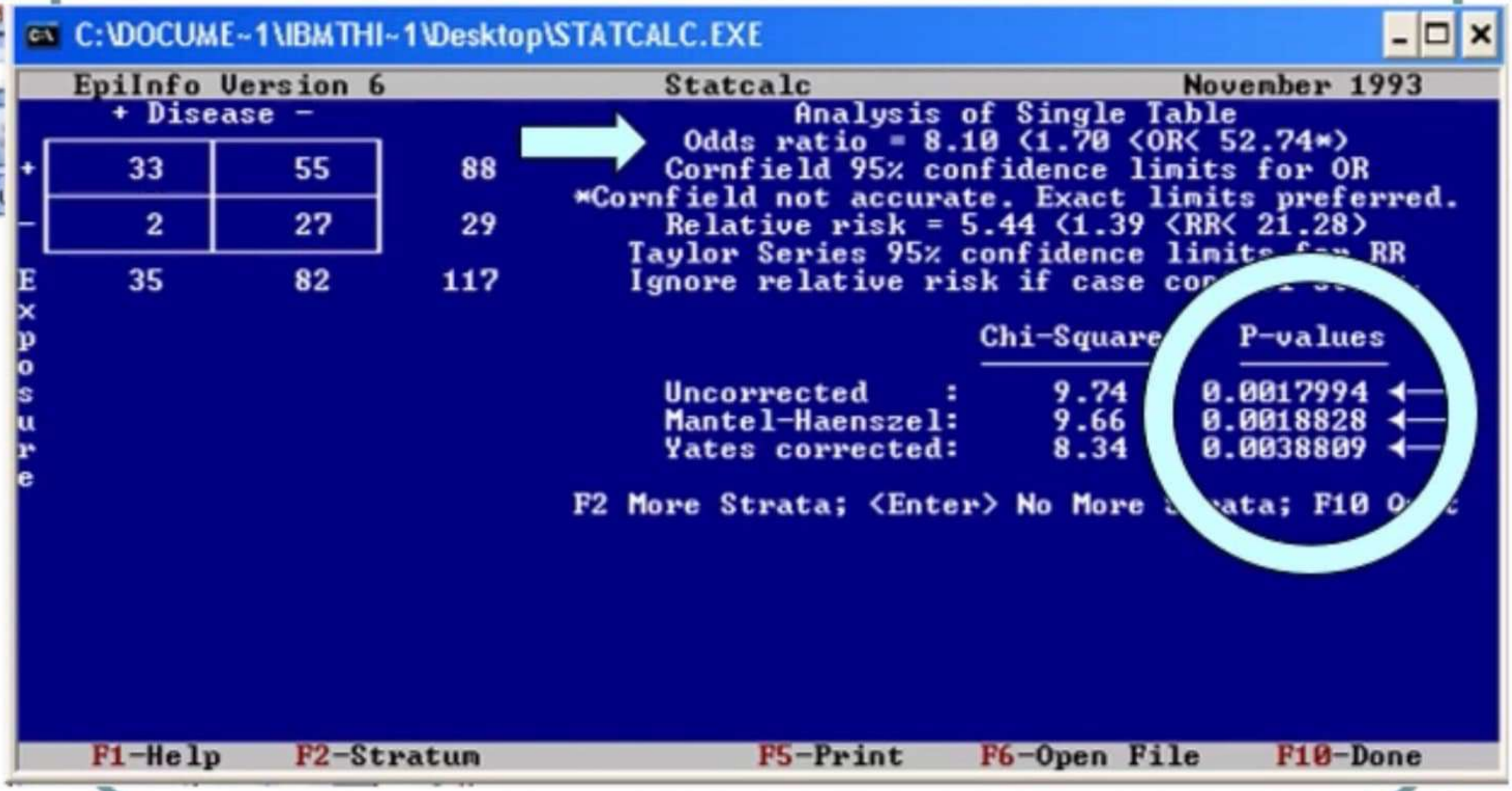
P- value

- We have found cigarette smokers has 8.1 times more risk of getting Lung cancer
- There are thousands of Lung cancer patients in the world
- We have taken only a small sample of 35 cases
- How do we know it is true for all lung cancer patients?

P-value

- To see if this association is due to chance.
- It is the probability that the difference is due to chance
- If P value is <0.05 it is considered statistically significant.
  - P value in lung cancer study is <0.001

Analysis


C:\DOCUME~1\IBMTHI~1\Desktop\STATCALC.EXE
EpiInfo Version 6
Statcalc
November 1993
+ Disease -
+ 33 55 88
- 2 27 29
35 82 117
Exposure
Analysis of Single Table
Odds ratio = 8.10 (1.70 <OR< 52.74*)
Cornfield 95% confidence limits for OR
*Cornfield not accurate. Exact limits preferred.
Relative risk = 5.44 (1.39 <RR< 21.28)
Taylor Series 95% confidence limits for RR
Ignore relative risk if case co
Chi-Square
P-values
Uncorrected : 9.74 0.0017994
Mantel-Haenszel: 9.66 0.0018828
Yates corrected: 8.34 0.0038809
F2 More Strata; <Enter> No More Strata; F10
F1-Help F2-Stratum F5-Print F6-Open File F10-Done

CC study - advantages

- Easy to conduct
- Inexpensive
- No risk to people
- No attrition (loss of patients) problems
- No ethical problems

CC study - disadvantages

- Problem of accuracy of data
  - Loss of memory
    - How many cigarettes a person smoked 20 years ago?
  - Incomplete records
    - What medicine a lady took in pregnancy?
- Getting good controls is difficult

Summary

- Case Control study is used to test hypothesis
- It involves four steps
  - Selection of cases and controls
  - Matching
  - Measuring exposure
  - Analysis (Exposure rate, Odds ratio and P value)

Summary

- The analysis of Case Control study is by a

  2x2 design

- Exposure rates are calculated among cases and controls

	Dis +	Dis -	
RF +	a	b	
RF -	c	d	
	a+c	b+d	

Summary

- Odds ratio is calculated to estimate the risk of disease among those who are exposed to the cause
- P value is calculated to know whether the difference is statistically significant

Cohort study

Cohort

- Is a group of people who share a common characteristic or experience
  - People born on a same day
  - Students who joined college in a year
  - People doing same work e.g. doctors

Cohort study

- Also called Prospective study or Incidence study.
- Is usually done **after** doing Case-Control study to get more proof of the cause of disease.
- The study is done on people **before** the disease occurs.

Design of Cohort study

	Disease +	Disease -	
Study cohort (Risk factor +)	a	b	a+b
Control cohort (Risk factor -)	c	d	c+d
	a+c	b+d	

Design of Cohort study

	Disease +	Disease -	
Study Cohort Risk factor +	a	b	a+b
Control Cohort Risk factor +	c	d	c+d
	a+c	b+d	

Elements of a Cohort study

1. Selection of study Cohort
2. Selection of comparison (Control) Cohort
3. Data collection and Follow-up
4. Analysis and interpretation

Selection of Study Cohort

- They are selected from general population or from specific groups e.g. Doctors, students etc.
- Members of the **study cohort** must **NOT** have the disease.
- Members of the **study cohort** must be exposed to the risk factor.

Selection of Control Cohort

- They are selected from general population or from specific groups e.g. Doctors, students etc.
- Members of the **control cohort** must **NOT** have the disease.
- Members of the **control cohort** must **NOT** be exposed to the risk factor.
- Members of the control cohort must be similar to the study cohort in age, sex etc.

Follow up

- Both the Study cohort and Control cohort is followed up to see how many develop the disease.
- This is done by
  - Medical examination
  - Personal visit, Phone call etc.
- Follow up is difficult because some persons will not respond.

Smoking and Lung cancer

	Disease + Lung cancer +	Disease – No Lung cancer	
Study Cohort Smoking +	70 (a)	6930 (b)	7000 a+b
Control Cohort Smoking -	3 (c)	2997 (d)	3000 c+d
	a+c	b+d	

	Disease + Lung cancer +	Disease – No Lung cancer	
Study Cohort Smoking +	70 (a)	6930 (b)	7000 a+b
Control Cohort Smoking -	3 (c)	2997 (d)	3000 c+d
	a+c	b+d	

Incidence rate among smokers = 70/7000 = 10 per 1000

Incidence rate among non-smokers = 3/3000 = 1 per 1000

Relative risk =

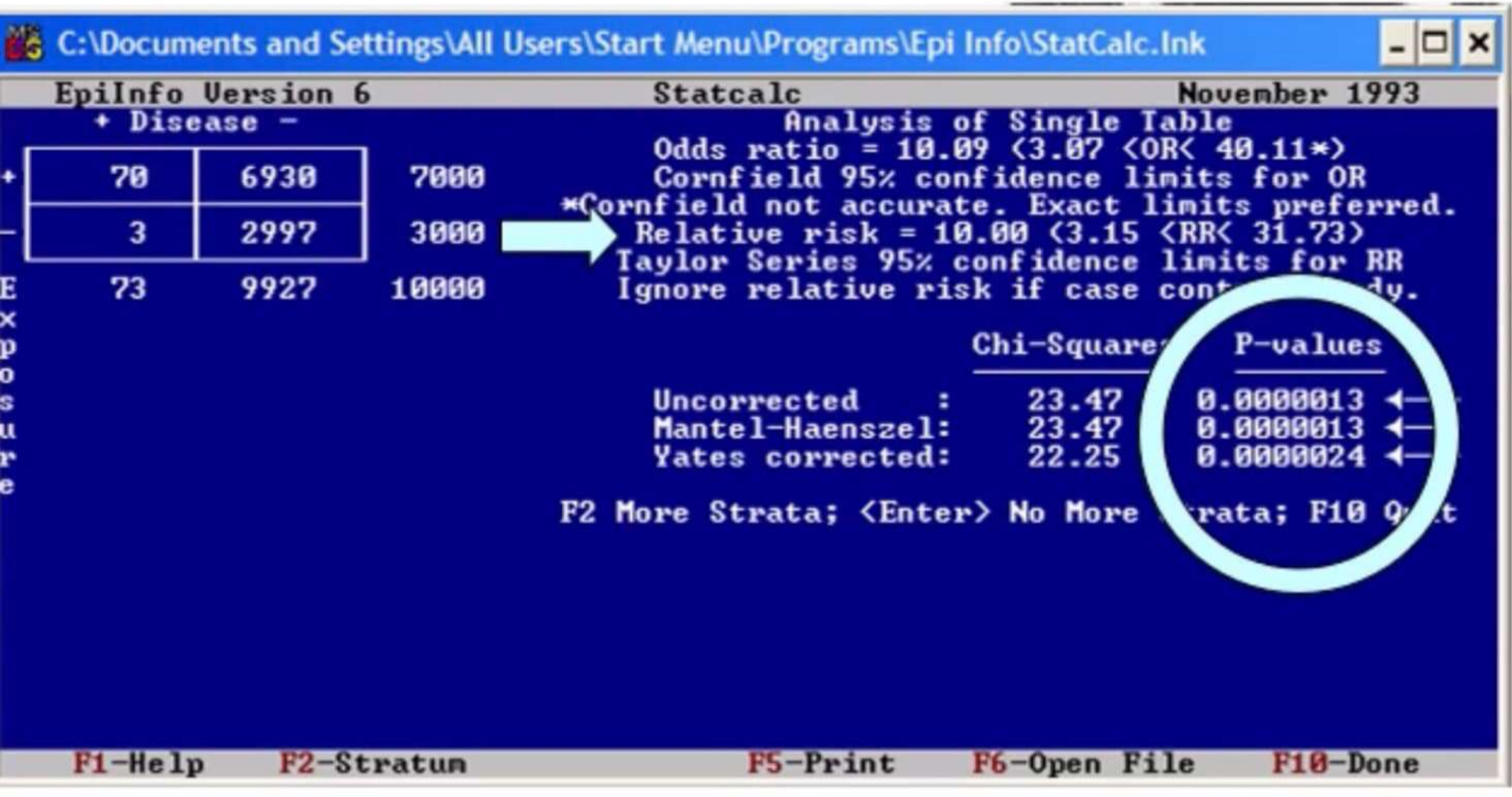
$$\frac{\text{Incidence of disease among exposed}}{\text{Incidence of disease among non-exposed}}$$

10/1= 10

- Relative risk of 10 means that those who smoke have 10 times more risk of developing lung cancer than those who are not smoking.
- If relative risk is 1, that means there is no risk.

P value

- There are many millions of cigarette smokers.
- We have studied only 7000.
- We have to know whether the Relative Risk of 10 is by chance or not.
- We calculate the P value. If the P value is less than(<) 0.05 we accept the Relative Risk.
- P value in our Lung cancer study is <0.001



This result is not by chance because P-value is < (less than) 0.05.

Advantage of Cohort study

- Incidence of disease can be calculated.
- More than one disease due to the risk factor can be studied.
  - Smoking and Lung cancer, peptic ulcer, Coronary heart disease etc.
- Gives better proof of the risk factor than Case Control study.

Disadvantages

- It takes long time to complete study.
- Persons may lose interest and will not come for follow-up.
- The person who is doing the study may lose interest or take another job.
- Cohort studies are expensive.
- Has more ethical problems.

Summary

- Cohort study gives better proof of the cause of disease.
- A group of people (Study Cohort) with the risk factor is selected.
- Another group of people (Control Cohort) without the risk factor is selected.
- Both groups are followed up to see how many develop disease.

Summary

- Incidence rate of disease is calculated among study cohort.
- Incidence rate of disease is calculated among the control cohort.
- Relative risk is calculated.
- Cohort study is more difficult and expensive than Case Control study.