



מכון גרטנר לחקר אפידמיולוגיה ומדיניות בריאות

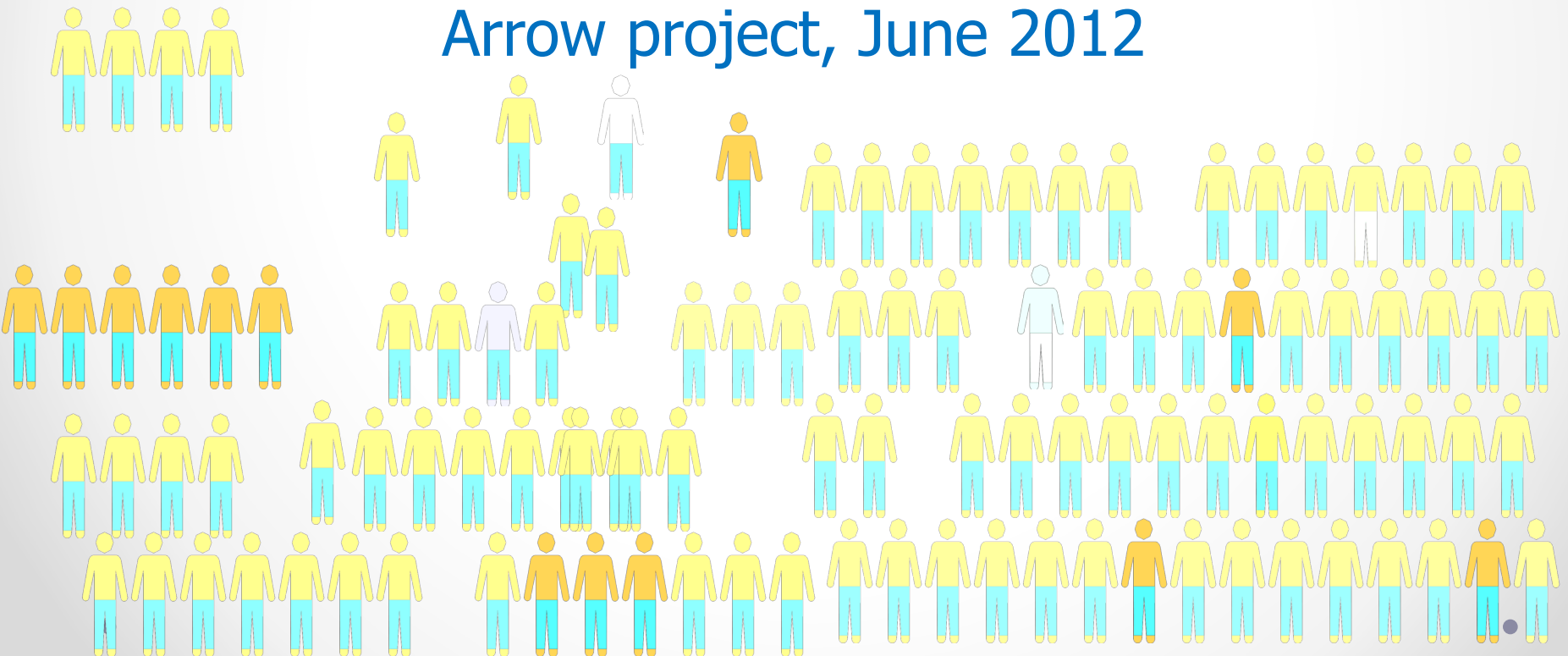
THE GERTNER INSTITUTE FOR EPIDEMIOLOGY AND HEALTH POLICY RESEARCH

Research Methods

Sample Size Calculation

Liat Lerner-Geva M.D., Ph.D

Arrow project, June 2012



STUDY DESIGN

```
graph TD; A[STUDY DESIGN] --> B[Analytic]; A --> C[Descriptive]; B --> D[Epidemiology Observational]; B --> E[Clinical Trial]; C --> F[Prevalence/ Incidence];
```

The diagram is a hierarchical flowchart. At the top is a dark blue rounded rectangle labeled 'STUDY DESIGN'. Two arrows point down from it to two light green rounded rectangles: 'Analytic' on the left and 'Descriptive' on the right. From 'Analytic', two arrows point down to 'Epidemiology Observational' (dark brown) and 'Clinical Trial' (dark grey). From 'Descriptive', one arrow points down to 'Prevalence/ Incidence' (tan). The boxes have colored borders: dark blue for the top, dark red for the middle row, dark green for 'Epidemiology Observational', and dark brown for 'Clinical Trial'.

Analytic

Descriptive

**Epidemiology
Observational**

Clinical Trial

**Prevalence/
Incidence**

Epidemiology Observational

**Exposure not
controlled**

**No-
randomization**

Clinical Trial

**Exposure
controlled by
investigator**

Randomization

Analytic Studies

□ □ □ □ -
□ □ □ □ □ □ □ □

**Exposed &
Outcome
Simultaneously**

RR



		Dis	
		+	-
Ex	+	a	b
	-	c	d

$$RR = \frac{a/a+b}{c/c+d}$$

□ □ □ □ -
□ □ □ □ □ □

**Cases Vs.
Controls**

OR



		Dis	
		+	-
Ex	+	a	b
	-	c	d

$$OR = \frac{a*d}{c*b}$$

□ □ □ □ □ □ □ □
□ □ □ □ □ □

**Exposed Vs.
Unexposed**

RR



		Dis	
		+	-
Ex	+	a	b
	-	c	d

$$RR = \frac{a/a+b}{c/c+d}$$

□ □ □ □ □ □
□ □ □ □ □ □ □ □

**Exposed Vs.
Unexposed**

RR



		Dis	
		+	-
Ex	+	a	b
	-	c	d

$$RR = \frac{a/a+b}{c/c+d}$$

Levels of Evidence

Box 2. Hierarchies of evidence for questions of therapy, prevention, aetiology or harm²⁶

- Level 1a** Systematic review (with homogeneity) of randomised controlled trials (RCTs)
- Level 1b** Individual RCT (with narrow confidence interval)
- Level 1c** All-or-none studies
- Level 2a** Systematic review (with homogeneity) of cohort studies
- Level 2b** Individual cohort study (including low quality RCT; eg <80% follow-up)
- Level 2c** 'Outcomes' research; ecological studies
- Level 3a** Systematic reviews (with homogeneity) of case-control studies
- Level 3b** Individual case-control study
- Level 4** Case series (and poor quality cohort and case-control studies)
- Level 5** Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

RR/OR > 1

Positive association- Risk factor

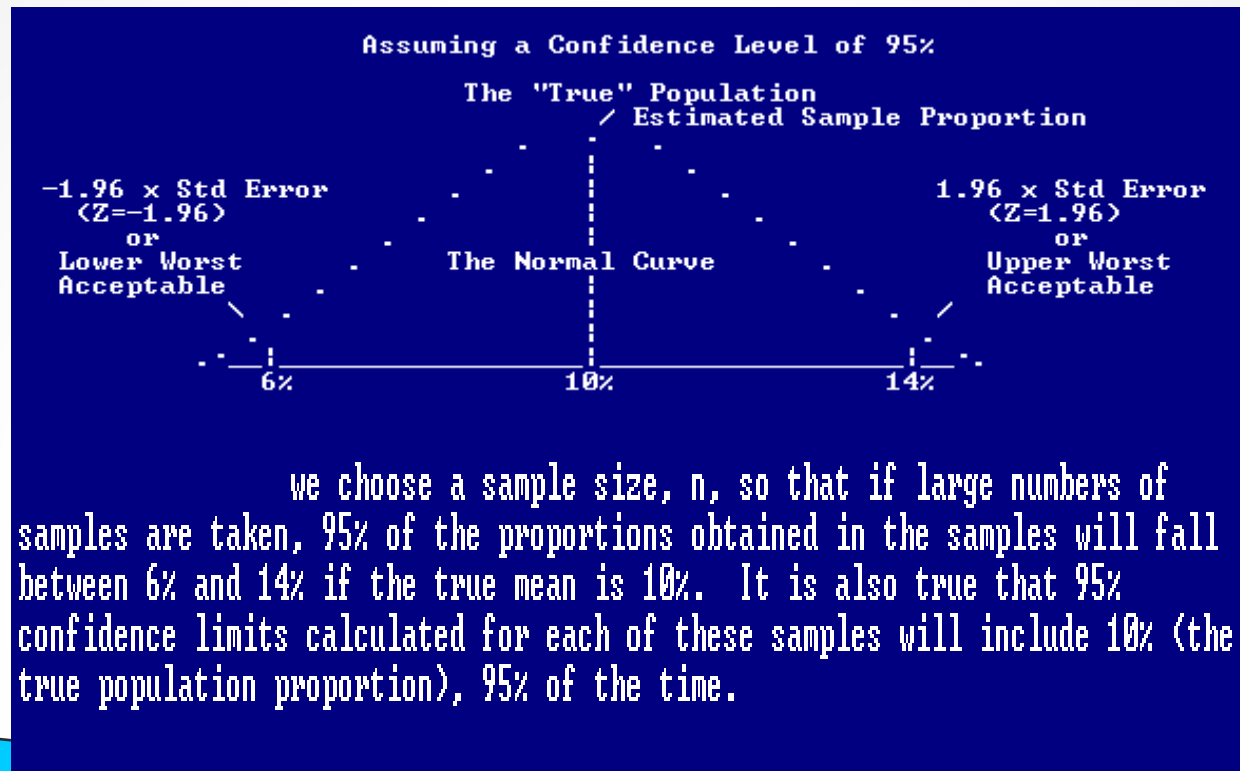
RR/OR = 1

No association

RR/OR < 1

Negative association- Protective factor

Confidential Interval



**95% CI
for P**

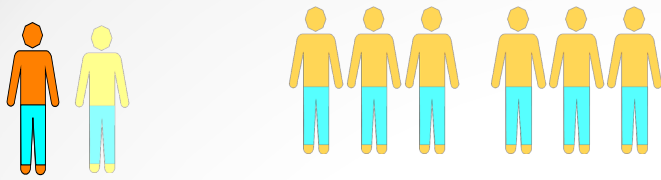
**A 95% confidence level means
that 95% of the intervals would
include the parameter;**

p

Sample

95%CI

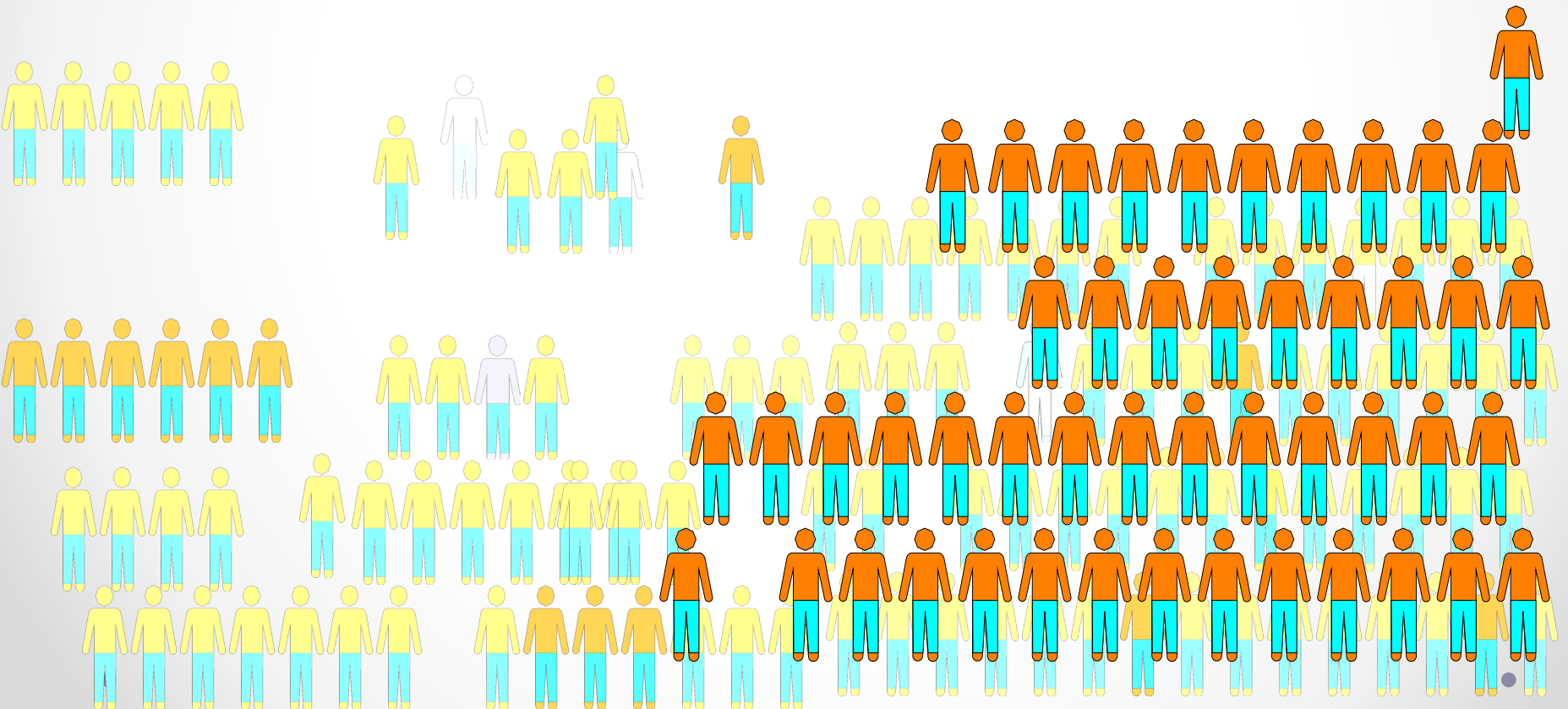
- **RR=1.2 95%CI 0.9-1.9**
- **RR=1.2 95%CI 1.1-18.7**
- **RR=1.2 95%CI 1.1-2.3**



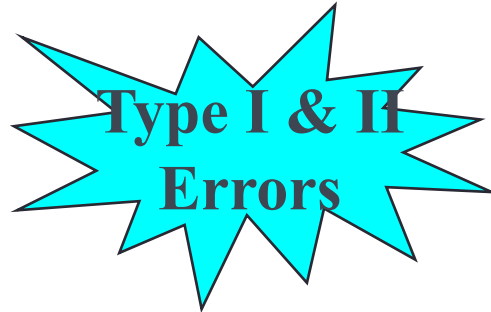
“How Many Subjects Do We Have To Study ?”

The time to answer this question is *before* the study is done.

Gordis L. Epidemiology 2nd edition. W.B Saunders Company, Philadelphia, 2000.

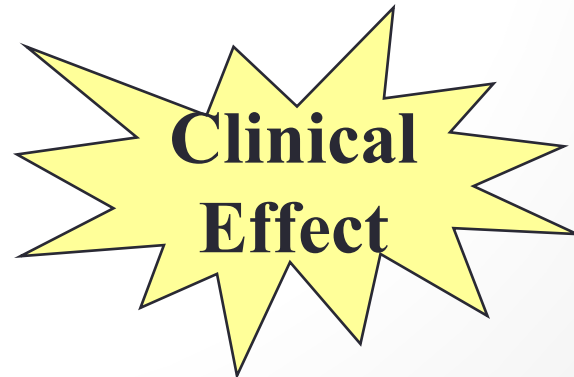


Sample Size



Type I error •
Type II error •

- **Prevalence**
- **Expected difference**



Type I Error / α

Probability that if the two **SAMPLES** differ this reflects a true difference in the two **POPULATIONS**

confidence level or $1 - \alpha$: **95%**

Type II Error/ β

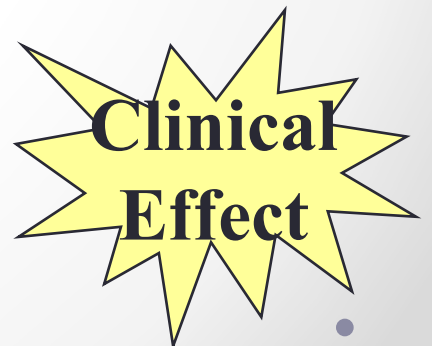
Power/ $1-\beta$



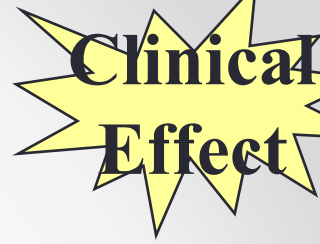
- No effect observed, when in fact there is a real effect
- If the test hypothesis is false, but is not rejected, the **incorrect** decision to **not reject** is called *type-II error*, or **beta error**.
- $POWER=1-\beta$

PREVALENCE & EXPECTED DIFFERENCES

- Literature
- Pilot studies
- Expert opinion
- Clinical significance
- Basic science evaluation
-



The effect of prevalence



$$RR=2$$

Frequent event 20%	Rare event 2%	
20/100 → 40/100	2/100 → 4/100	Small sample
200/1000 → 400/1000	20/1000 → 40/1000	Large sample

WWW.CDC.EPIINFO

CDC Home



Centers for Disease Control and Prevention

CDC 24/7: Saving Lives. Protecting People.™

SEARCH

A-Z Index [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) <#>

Epi Info™

Epi Info™

[Introducing Epi Info™ 7](#)

[Downloads](#)

[Translations](#)

[Training Resources](#)

[Vendors](#)

[Shapefiles](#)

[Help Desk](#)

[Frequently Asked Questions](#)

Related Links

[Epi Info™ /Mesh4x](#)

[Epi Info™ Community](#)

[Open Epi](#)

Introducing...



What is Epi Info™?

Physicians, nurses, epidemiologists, and other public health workers lacking a background in information technology often have a need for simple tools that allow the rapid creation of data collection instruments and data analysis, visualization, and reporting using epidemiologic methods. Epi Info™, a suite of lightweight software tools, delivers core ad-hoc epidemiologic functionality without the complexity or expense of large, enterprise applications.

Epi Info™ is easily used in places with limited network connectivity or limited resources for commercial software and professional IT support. Epi Info™ is flexible, scalable, and free while enabling data collection, advanced statistical analyses, and geographic information system (GIS) mapping capability.


Since its initial release, Epi Info™ users have self-registered in over 181 countries covering all continents including Antarctica. Epi Info™ has been translated in more than 13 languages.


More than one million users are estimated.

[Email page](#)

[Print page](#)

Contact Us:

 Epi Info™ Help Desk
Centers for Disease
Control and
Prevention
1600 Clifton Rd
Mail Stop E-91
Atlanta, GA 30333

 **404-498-6190**
Monday-Friday, 8:30
a.m. - 4:30 p.m.
(Eastern Standard
Time)

 epiinfo@cdc.gov

COHORT

(power = 80%, $\alpha=0.05$)

Tables (2 x 2, 2 x n)

Sample size & power

Ch

Population survey

Cohort or cross-sectional

Unmatched case-control

Statcalc

EpiInfo Version 6

Statcalc

November 1993

Unmatched Cohort and Cross-Sectional Studies (Exposed and Nonexposed)

Probability that if the two SAMPLES differ this reflects a true difference in the two POPULATIONS (Confidence level or 1- α)

: 25.00 %

Probability that if the two POPULATIONS differ, the two SAMPLES will show a "significant" difference (Power or 1- β)

: 80.00 %

Ratio (Number of Unexposed : Number of Exposed) :

1 : 1

Expected frequency of disease in unexposed group

: 1.00 %

Please fill in the closest value to be detected for ONE of the following:

Risk ratio (RR) or relative risk--closest to 1.00

: 1.10

Odds ratio (OR)--closest to 1.00

: 1.10

Percent disease among exposed--closest to % for unexposed

: 1.10 %

F1-Help

F4-Calc

F6-Open File

F10-Done

Unmatched Cohort and Cross-Sectional Studies (Exposed and Nonexposed)
Sample Sizes for 1.00 % Disease in Unexposed Group

Conf.	Power	Unex:Exp	Disease in Exposed	Risk Ratio	Odds Ratio	Sample Size Unexp.	Exposed	Total
95.00 %	80.00 %	1:1	1.10 %	1.10	1.10	165,089	165,089	330,178
90.00 %	"	"				130,462	130,462	260,924
95.00 %	"	"				165,089	165,089	330,178
99.00 %	"	"				244,678	244,678	489,356
99.90 %	"	"				356,798	356,798	713,596
95.00 %	80.00 %	"				165,089	165,089	330,178
"	90.00 %	"				220,332	220,332	440,664
"	95.00 %	"				272,017	272,017	544,034
"	99.00 %	"				383,763	383,763	767,526
"	80.00 %	4:1				408,048	102,012	510,060
"	"	3:1				327,069	109,023	436,092
"	"	2:1				246,086	123,043	369,129
"	"	1:2				124,576	249,153	373,729
"	"	1:3				111,068	333,205	444,273
"	"	1:4				104,313	417,252	521,565

Change values for
inputs as desired,
then press F4 to
recalculate.

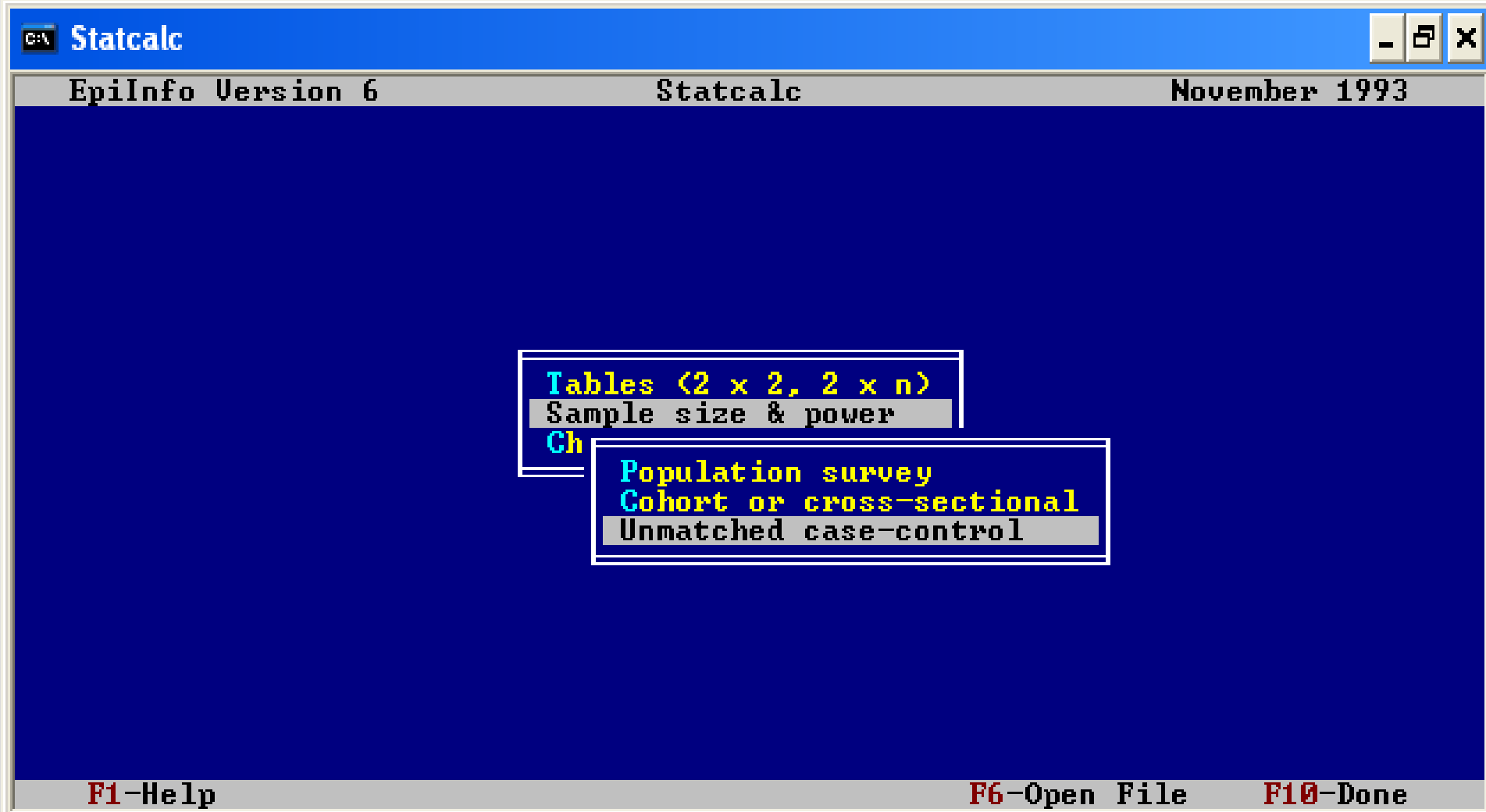
Cohort Study-Examples

(power = 80%, $\alpha=5\%$)

Unexposed (n)	Exposed (n)	Outcome among Unexposed	RR- minimal	Unexposed/ Exposed Ratio
165,089	165,089	1.0	1.1	1.0
16,389	16,389	0.5	1.5	1.0
8,145	8,145	1.0	1.5	1.0
5,065	5,065	0.5	2.0	1.0

Case-Control Study

(power = 80%, CI=95%)



Unmatched Case-Control Study (Comparison of ILL and NOT ILL)

Probability that if the two SAMPLES differ this reflects a true difference in the two POPULATIONS (Confidence level or $1-\alpha$) : 95.00 %

Probability that if the two POPULATIONS differ, the two SAMPLES will show a "significant" difference (Power or $1-\beta$) : 80.00 %

#NOT ILL/#ILL (1 means equal sample sizes) 1 : 1

Expected frequency of exposure in NOT ILL group : 15.00 %

Please fill in the closest value to be detected for ONE of the following:

Odds ratio (OR)--closest to 1.00 : 2.00

Percent exposure among ILL group--closest to % for NOT ILL : 26.09 %

Unmatched Case-Control Study (Comparison of ILL and NOT ILL)
Sample Sizes for 15.00 % Exposure in NOT ILL Group

Conf.	Power	NOT ILL :ILL	Exposure in ILL	Odds Ratio	NOT ILL	Sample Size
95.00 %	80.00 %	1:1	26.09 %	2.00	225	2
90.00 %	"	"			181	1
95.00 %	"	"			225	2
99.00 %	"	"			327	3
99.90 %	"	"			470	4
95.00 %	80.00 %	"			225	2
"	90.00 %	"			295	2
"	95.00 %	"			360	3
"	99.00 %	"			501	5
"	80.00 %	4:1			532	1
"	"	3:1			432	1
"	"	2:1			328	1
"	"	1:2			172	3
"	"	1:3			155	4
"	"	1:4			146	5

Change values for
inputs as desired,
then press F4 to
recalculate.

Case-Control Study

(power = 80%, CI=95%)

Controls (n)	Cases (n)	% Exposed in Controls	Minimal OR	Cases/Controls Ratio
225	225	15.0	2.0	1.0
691	691	15.0	1.5	1.0
447	447	30.0	1.5	1.0
676	338	30.0	1.5	2.0

In summary....

Sample Size	
↓	↑ Prevalence
↓	↑ Expected RR / OR

