

2019 – 2020

Dr. Fazıl Küçük Faculty of Medicine, EMU  
Year 1 Committee 3

# Biostatistics Course

Instructor: Assist. Prof. Dr. İlke Akçay  
[ilke.akcay@emu.edu.tr](mailto:ilke.akcay@emu.edu.tr)

# Covered Lectures in Y1C1

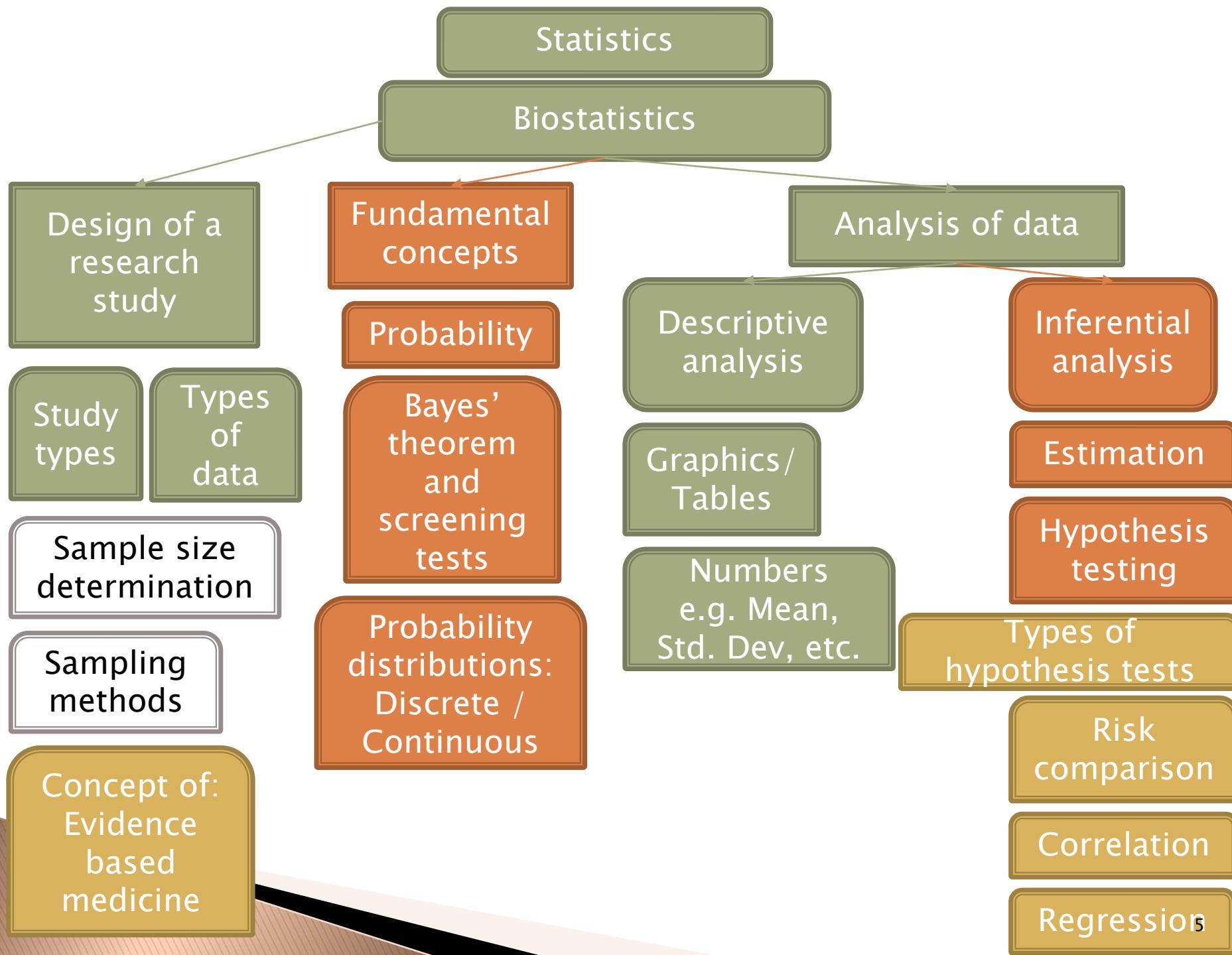
- ▶ What is statistics and biostatistics?
- ▶ Statistics in medical research
- ▶ Designing research (Study types)
- ▶ Types of data
- ▶ Describing data with graphics
- ▶ Describing data with numbers

# Covered Lectures in Y1C2

- ▶ What is probability and probability distribution?
- ▶ Bayes' Theorem
- ▶ Principles of statistical analysis
- ▶ Elements of statistical inference
- ▶ Introduction to statistical analysis
- ▶ Sampling, distribution and estimation
- ▶ Testing statistical hypothesis
- ▶ Types of errors in statistical inference
- ▶ Difference between parametric and nonparametric methods; Introduction to parametric methods
- ▶ One sample t-test, unpaired t-test and paired t-test

# Topics of Y1C3 biostatistics course

- ▶ Review of statistical inference table and introduction to nonparametric methods
- ▶ Sign test, Mann-Whitney U test, Wilcoxon test (paired sample)
- ▶ Chi-square test of independence
- ▶ Chi-square test of homogeneity
- ▶ Comparing risk: Relative Risk and Odds ratio
- ▶ Some basic concepts for diagnosis tests
- ▶ Correlation
- ▶ Simple Regression
- ▶ Approach to Evidence-based Medicine
- ▶ Statistical packages
- ▶ Interpretation of computer outputs
- ▶ Systematic reviews and Meta analysis



# Lecture 1 of Y1 C3

Review of statistical inference table and introduction to  
nonparametric methods

Sign test, Mann–Whitney U test, Wilcoxon test (paired  
sample)

# Review of Statistical Inference

## Statistical Inference

### Estimation

- Point Estimate
- Interval Estimate

### Hypothesis Testing

- Parametric Methods
- Nonparametric Methods

# Recall

## Steps in Hypothesis Testing

- ▶ 1. Evaluate data
- ▶ 2. Review assumptions
  - parameters related to the population, distribution
- ▶ 3. State hypotheses
- ▶ 4A. Select and calculate test statistics
- ▶ 4B. Find p-value
- ▶ 5. State decision rule
- ▶ 6. Make statistical decision:
  - Reject  $H_0$  (conclude  $H_1$  is true)
  - Fail to reject  $H_0$  (conclude  $H_0$  might be true)
- ▶ 7. Conclusion

# Parametric vs. Nonparametric test procedures

## Parametric

- ▶ Assume that data are normally distributed
- ▶ Require interval scale or ratio scale
- ▶ More complex computations
- ▶ Tend to have less power when there are extreme values (outliers)
- ▶ Make inferences about mean(s)

## Nonparametric

- ▶ There is no assumption on the distribution of the data
- ▶ Data measured on any scale
- ▶ Simpler computations
- ▶ Can be useful for dealing with extreme values (outliers)
- ▶ Make inferences about median(s)

# How to determine which hypothesis test to use?

- ▶ When deciding on the appropriate hypothesis test, we should consider:
  - 1) The type of the data
  - 2) The distribution of the data
  - 3) The number of groups of observations
  - 4) If there are groups, the independence/dependence of the group of observations

		<b>Parametric Methods</b>	<b>Non parametric Methods</b>
		<b>Assumption:</b> Data have to be sampled from Gaussian distribution (normally distributed)&populations have equal variances (homogeneity of variances)	No assumption. It is preferred when normality is violated and having small sample size ( $n < 30$ )
Quantitative Data	<b>1 sample vs hypothesized value comparison</b>	One-sample t-test	Signed rank test or Wilcoxon Rank Sum Test
	<b>2 samples Comparisons</b>	Unpaired t-test	Mann-Whitney U Test
		Paired t-test	Wilcoxon Matched Pairs Test
	<b>More than 2 samples Comparisons</b>	Independent Samples  One way-ANOVA Test If $p < 0.05$ according to F-test, Multiple Comparison Tests are needed; such as; Tukey, Bonferroni, Scheffé Multiple Comparison Tests To compare all vs. only with control group, Dunnett Multiple Comparison Test can be used.	Kruskal-Wallis Test If $p < 0.05$ use Dunn Multiple Comparison Test
		Paired Samples or Matched Pair Samples  Repeated Measures ANOVA Test If $p < 0.05$ according to F-test, Multiple Comparison Tests are neededStudent such as Newman-Keuls	Friedman Test If $p < 0.05$ use Dunn Multiple Comparison Test
	Crosstabulation	Independent Samples  ---	<b>X<sup>2</sup>-Test</b> * Homogeneity test *Independence test If the association is significant ; Phi or Cramer's V coefficients can be used In order to measure the degree and to determine the direction of the association.
Categorical (qualitative) Data		Paired Samples or Matched Pair Samples  ---	McNemar X <sup>2</sup> test (2x2) or Stuart-Maxwell test(3x3)
<b>Correlation Coefficient</b> (This measures the degree and determines the direction of linear co-relation between two continuous variables, x and y)		Pearson Correlation Coefficient	Spearman Correlation Coefficient

# Recall t-tests

## 1) One sample t-test

- We are interested in whether the mean of the variable in the population of interest differ from a specific hypothesized mean.
- Comparison parameter has been estimated from prior research or is derived from theory.

## 2) Independent samples (or Unpaired) t-test

- We compare the means of two independent populations.

## 3) Paired samples t-test

- We compare two dependent groups.

# Example: One Sample t-test

- ▶ A researcher is planning a psychological intervention study, but before he proceeds he wants to characterize his participants' depression levels. He tests each participant on a particular depression index, where anyone who achieves a score of 4.0 is deemed to have 'normal' levels of depression. Lower scores indicate less depression and higher scores indicate greater depression. He has recruited 40 participants to take part in the study. Depression scores are recorded in the variable dep\_score. He wants to know whether his sample is representative of the normal population (i.e., do they score statistically significantly differently from 4.0).
- ▶ Data: 3.68, 3.98, 3.72, 3.98, ... , 4.26, 3.51, 4.04 (n=40)

# Example: One Sample t-test

- ▶ The example is a one sample design to compare the population mean with a known threshold.
- ▶ Can we directly apply one-sample t-test?
  - No!
  - First, we need to check if the assumption of normality is met.

# Normality check

- In order to check if the data is normally distributed or not, Shapiro-Wilk test is conducted.
  - $H_0$ : Data is normally distributed
  - $H_1$ : Data is not normally distributed

Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
dep_score	.103	40	.200*	.958	40	.138

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

P-value=0.138 > 0.05

We fail to reject  $H_0$

Data is normally distributed

# Example: One Sample t-test

- ▶ Since data is normally distributed, one sample t-test can be conducted:
  - $H_0$ : The mean depression score of the study population is not different than 4 ( $\mu_{dep\_score} = 4$ )
  - $H_1$ : The mean depression score of the study population is different than 4 ( $\mu_{dep\_score} \neq 4$ )

# Example: One Sample t-test

## ► Results:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
dep_score	40	3.5158	.58434	.09239

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
dep_score	-5.241	39	.000	-.48425	-.6711	-.2974

P-value=0.000 < 0.05

We reject  $H_0$

$\mu_{dep\_score} \neq 4$ , it is less than 4

Lower by 0.48

CI of the difference is (-0.67,-0.3)

# Example: One Sample t-test

- ▶ How to report these results?

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
dep_score	40	3.5158	.58434	.09239

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
dep_score	-5.241	39	.000	-.48425	-.6711	-.2974

- ▶ Mean depression score ( $3.52 \pm 0.58$ ) was lower than the normal depression score of 4.0, with a statistically significant difference of 0.48 (95% CI, 0.3 to 0.67),  $t(39) = -5.241$ ,  $p < 0.0005$ .

# Example: Independent Samples t-test

- The table below shows the cholesterol levels of 20 patients from City A, and 18 patients from City B. Are the means of cholesterol levels from these two cities statistically different from each other?

	cholesterol levels (mmol/l)									
	1	2	3	...	17	18	19	20		
city A	4.1	5.3	4.9	...	5.2	6.4	5.1	5.8		(n=20)
city B	6.6	5.2	5.9	...	6.3	6.4				(n=18)

# Example: Independent Samples t-test

- ▶ The example is a two independent samples design to compare the population means.
- ▶ Can we directly apply independent samples t-test?
  - No!
  - First, we need to check the assumptions
    - Observations in each group should follow a normal distribution
      - Shapiro-Wilk test
    - The variances in the two samples should be equal (homogeneous)
      - Levene's test for homogeneity of variances (this test is done within the procedure of t-test / so no extra work needed)

# Normality check

- In order to check if the data in groups is normally distributed or not, Shapiro-Wilk test is conducted.
  - $H_0$ : Data is normally distributed
  - $H_1$ : Data is not normally distributed

Tests of Normality							
group	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
Cholesterol (mmol/l)	City A	.117	20	.200*	.973	20	.815
	City B	.165	18	.200*	.932	18	.215

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

For City A: p-value=0.815 > 0.05

We fail to reject  $H_0$  ; data from City A is normal

For City B: p-value=0.215 > 0.05

We fail to reject  $H_0$  ; data from City B is normal

# Example: Independent Samples t-test

- ▶ Since data is normally distributed, independent samples t-test can be conducted:
  - $H_0$ : The mean cholesterol levels in City A is not different than the mean cholesterol level in City B ( $\mu_{City\ A} = \mu_{City\ B}$ )
  - $H_1$ : The mean cholesterol levels in City A is different than the mean cholesterol level in City B ( $\mu_{City\ A} \neq \mu_{City\ B}$ )

# Example: Independent Samples t-test

## ► Results:

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
Cholesterol (mmol/l)	City A	20	5.670	.7740	.1731
	City B	18	6.006	.6958	.1640

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	Lower	Upper
Cholesterol (mmol/l)	Equal variances assumed	.522	.475	-1.399	36	.170	-.3356	.2398	-.8219	.1508
	Equal variances not assumed			-1.407	36.000	.168	-.3356	.2384	-.8191	.1480

### Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference			
								Lower	Upper	
Cholesterol (mmol/l)	Equal variances assumed	.522	.475	-1.399	36	.170	.3356	.2398	-.8219	.1508
	Equal variances not assumed			-1.407	36.000	.168	.3356	.2384	-.8191	.1480

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$p = 0.475 > 0.05$$

Fail to reject  $H_0$   
Variances are equal

$$H_0: \mu_{City A} = \mu_{City B}$$

$$H_1: \mu_{City A} \neq \mu_{City B}$$

$$p = 0.17 > 0.05$$

Fail to reject  $H_0$   
Means are not different

Mean cholesterol level in City A( $5.67 \pm 0.77$ ) does not significantly differ from the mean cholesterol level in City B ( $6 \pm 0.7$ ),  $t(36) = -1.399$ ,  $p = 0.17$ .

# Example:

## Paired Samples t-test

- In a tumor size study, two doctors were shown the same set of tumor pictures. The volume of tumor was measured (in cm<sup>3</sup>) by two separate physicians under similar conditions, and the following values are obtained. Is there a difference in tumor volume measurement based on physician?

Tumor	Dr1	Dr2
1	15.8	17.2
2	22.3	20.3
3	14.5	14.2
4	15.7	18.5
5	26.8	28.0
6	24.0	24.8
7	21.8	20.3
8	23.0	25.4
9	29.3	27.5
10	20.5	19.7

# Example:

## Paired Samples t-test

- ▶ The example is a paired samples design to compare the population means
- ▶ Can we directly apply paired samples t-test?
  - No!
  - First, we need to check the assumption of normality:
    - The differences should be normally distributed
      - Shapiro-Wilk test

# Normality check

- In order to check if the data is normally distributed or not, Shapiro-Wilk test is conducted.
  - $H_0$ : Data is normally distributed
  - $H_1$ : Data is not normally distributed

Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
difference	.138	10	.200*	.930	10	.452

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

$$\text{P-value} = 0.452 > 0.05$$

We fail to reject  $H_0$

Data is normally distributed

# Example:

## Paired Samples t-test

- ▶ Since data is normally distributed, paired samples t-test can be conducted:
  - $H_0$ : There is no difference in tumor volume measurement based on physician  
 $(\mu_{dr1} = \mu_{dr2})$
  - $H_1$ : There is a difference in tumor volume measurement based on physician  
 $(\mu_{dr1} \neq \mu_{dr2})$

# Example: Paired Samples t-test

## ► Results:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	dr1	21.3700	10	4.87762	1.54244
	dr2	21.5900	10	4.60880	1.45743

High positive correlation  
( $r = 0.934, p < 0.0005$ ),

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 dr1 & dr2	10	.934	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1 dr1 - dr2	-.22000	1.74407	.55152	-1.46763	1.02763	-.399	9	.699			

### Paired Samples Test

	Paired Differences						t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference								
				Lower	Upper							
Pair 1 dr1 - dr2	-0.22000	1.74407	.55152	-1.46763	1.02763	-0.399	9	0.699				

We can conclude that mean volume measurement by physician1  $21.37 \pm 4.88$  is not significantly different than the mean volume measurement by physician2  $21.59 \pm 4.61$ ;  $t(9) = -0.399$ ;  $p = 0.699$

$$\begin{aligned} H_0: \mu_{dr1} &= \mu_{dr2} \\ H_1: \mu_{dr1} &\neq \mu_{dr2} \end{aligned}$$

$$p = 0.699 > 0.05$$

Fail to reject  $H_0$   
Means are not different

# Nonparametric Tests for Quantitative Data

- ▶ What if data is not normally distributed and t-tests cannot be conducted?

Type of Design	Parametric Test	Nonparametric Test
One sample	One-sample t-test	Sign Test
Two independent samples	Independent-samples t-test	Mann Whitney U test
Two paired sample	Paired-samples t-test	Wilcoxon Signed Ranks test
.	.	.
.	.	.
.	.	.

# One Sample Case

Example: For the 21 patients with medullary carcinoma, percentages of cells with aberrations were calculated. Suppose that, in general population, the mean percentage of chromosomal aberrations has been reliably estimated to be 3.90. Does the population of patients with medullary carcinoma differ from the general population in the percentage of cells with chromosomal aberrations?

Subject	Percentages of cells with aberrations	Subject	Percentages of cells with aberrations
1	13.9	12	5.0
2	10.9	13	2.0
3	9.3	14	2.0
4	5.2	15	17.1
5	14.0	16	13.4
6	2.0	17	6.0
7	3.2	18	14.0
8	12.0	19	18.0
9	13.0	20	13.6
10	11.3	21	5.4
11	12.2		

Example: Doctor Z et al. reported the endurance scores of 12 animals during a 48-hour session of discrimination responding. Conduct a test to see whether the investigators may conclude at the 0.05 level of significance that the median endurance score of animals with electrodes implanted in the forebrain is equal to 97.5 or not.

Subject	Lead Level
1	98.1
2	82.4
3	97.7
4	84.4
5	97.8
6	94.5
7	81.7
8	97.5
9	80.3
10	94.6
11	85.5
12	82.6

# One Sample Case

## Example:

- 21 patients
- the mean percentage in general population is estimated to be 3.90.
- Does the population of patients with medullary carcinoma differ from the general population by means of the percentage of cells with chromosomal aberrations?

Shapiro-Wilk Test of Normality ->  
p=0.057

## Parametric: One sample t-test

### Hypotheses

$$\begin{aligned} H_0: \mu &= 3.90 \\ H_1: \mu &\neq 3.90 \end{aligned}$$

Sample mean = 9.69  
Sample Std Dev = 5.14  
Mean Difference = 5.79  
**p<0.0005 ; Reject  $H_0$**

Conclusion: Mean percentage of cells with chromosomal aberrations in the population of patients with MC is significantly different (higher) than the general population.

## Example:

- the endurance scores of 12 animals during a 48-hour session of discrimination responding.
- test whether the investigators may conclude that the median endurance score of animals with electrodes implanted in the forebrain is equal to 97.5 or not.

Shapiro-Wilk Test of Normality ->  
p= 0.013

## Nonparametric: One sample signed rank test

### Hypotheses

$$\begin{aligned} H_0: \text{The median in the population equals } 97.5. \\ H_1: \text{The median in the population does not equal } 97.5. \end{aligned}$$

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of lead_level equals 97.500.	One-Sample Wilcoxon Signed Rank Test	.016	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Conclusion: The median endurance score of animals with electrodes implanted in the forebrain significantly differs from 97.5

# Two Independent Samples

Example: The following data from Schechter et al. [1973] deal with sodium chloride preference as related to hypertension. Two groups, 12 normal and 10 hypertensive subjects, were isolated for a week and compared with respect to Na<sup>+</sup> intake. Data includes the average daily Na<sup>+</sup> intakes (in milligrams). Compare the average daily Na<sup>+</sup> intake of the hypertensive subjects with that of the normal volunteers by means of an appropriate hypothesis test.

Average daily Na <sup>+</sup> intakes		
Subjects	Hypertensive Group (n=10)	Control Group (n=12)
1	92.8	39.7
2	54.8	24.1
3	51.6	66.3
4	61.7	22.5
5	52.4	47.1
6	84.5	42.2
7	42.1	39.3
8	62.2	43.6
9	43.7	51.3
10	55.6	44.3
11		43.1
12		39.8

Example: A company is interested in marketing a clotting agent that reduces blood loss when an accident causes an internal injury such as liver trauma. The company conducts an experiment in which a controlled liver injury is induced in pigs and blood loss is measured. Pigs are randomized as to whether they receive the drug after the injury or do not receive drug therapy—the treatment and control groups, respectively. The data were taken from a study in which there were 10 pigs in the treatment group and 10 in the control group. The blood loss was measured in milliliters. We want to conclude whether blood loss in two groups differs or not.

Blood Loss in milliliters		
Subjects	Treatment Group (n=10)	Control Group (n=10)
1	786	543
2	375	666
3	4446	455
4	2886	823
5	478	1716
6	587	797
7	434	2828
8	4764	1251
9	3281	702
10	3837	1078

# Two Independent Samples

## Shapiro-Wilk Test of Normality

Tests of Normality

group	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
daily Na+ intakes	.250	10	.076	.864	10	.085
hypertensive control	.242	12	.051	.904	12	.180

a. Lilliefors Significance Correction

## Parametric: Independent samples t-test

### Hypotheses

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

### Group Statistics

group		N	Mean	Std. Deviation	Std. Error Mean
daily Na+ intakes	hypertensive	10	60.140	16.4804	5.2116
	control	12	41.942	11.4023	3.2916

### Independent Samples Test

daily Na+ intakes	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
								Lower Upper
Equal variances assumed	1.403	.250	3.054	20	.006	18.1983	5.9596	5.7668 30.6299
Equal variances not assumed			2.952	15.583	.010	18.1983	6.1640	5.1028 31.2939

p=0.006 ; Reject  $H_0$ 

Conclusion: There is a significant difference between two groups by means of their daily Na+ intake.  
 Daily Na+ intake in hypertensive patients is higher than the daily Na+ intake in control group.

## Shapiro-Wilk Test of Normality

Tests of Normality

group	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
blood_loss	.279	10	.027	.826	10	.030
treatment	.243	10	.097	.796	10	.013

a. Lilliefors Significance Correction

## Nonparametric: Mann Whitney U test

### Hypotheses

$H_0$ : There is no difference between treatment and control groups by means of the median of blood loss.  
 $H_1$ : There is a difference between treatment and control groups by means of the median of blood loss.

### Test Statistics<sup>b</sup>

	blood_loss
Mann-Whitney U	43.000
Wilcoxon W	98.000
Z	-.529
Asymp. Sig. (2-tailed)	.597
Exact Sig. [2*(1-tailed Sig.)]	.631 <sup>a</sup>

a. Not corrected for ties.

b. Grouping Variable: group

p=0.597 ; Fail to reject  $H_0$ 

Conclusion: There is no significant difference between treatment and control groups with respect to their blood loss.

# Paired Samples

Example: A random sample of 10 young men was taken and the heart rate (HR) of each young man was measured before and after having a cup of caffeinated coffee. The results were given (beats / min). Does caffeinated coffee have any effect on the heart rate of young men ?

Heart rates (beats per minute)		
Subjects	HR before coffee	HR after coffee
1	68	74
2	64	68
3	52	60
4	76	72
5	78	76
6	62	68
7	66	72
8	76	76
9	78	80
10	60	64

Example: Weis and Peak studied the effects of oxytocin on blood pressure during anesthesia. The subjects were 11 women, 19 to 31 years of age and weighted 103 to 251 pounds and were in the first trimester of pregnancy. They had been anesthetized for dilation and curettage, and were injected with 0.1 unit/kg of oxytocin. The arterial blood pressures before and after oxytocin are recorded. Can we conclude on the basis of these data that oxytocin lowers the arterial blood pressure of women?

Arterial Blood Pressures (mmHg)		
Subjects	Before	After
1	95	92
2	173	90
3	94	80
4	97	59
5	81	72
6	100	46
7	97	75
8	104	92
9	72	70
10	101	34
11	83	78

# Paired Samples

## Shapiro-Wilk Test of Normality

Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
difference	.201	10	.200*	.924	10	.394

## Parametric: Paired t-test

### Hypotheses

$H_0$ : The mean heart rate of young man does not differ after having a cup of caffeinated coffee.  
 $(\mu_{HR\_before} = \mu_{HR\_after})$

$H_1$ : The mean heart rate of young man differs after having a cup of caffeinated coffee.  
 $(\mu_{HR\_before} \neq \mu_{HR\_after})$

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 HRbefore	68.00	10	8.844	2.797
HRafter	71.00	10	6.055	1.915

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 HRbefore & HRAfter	10	.929	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1 HRbefore - HRAfter	-3.000	3.916	1.238	-5.801	-.199	-2.423	9	.038			

p=0.038 ; Reject  $H_0$

Conclusion: The mean heart rate significantly increases after coffee.

## Shapiro-Wilk Test of Normality

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
ABPafter_ABPbefore	.851	11	.044

## Nonparametric: Wilcoxon matched pairs test

### Hypotheses

$H_0$ : The median of Arterial Blood Pressure does not differ after oxytocin is applied.

$H_1$ : The median of Arterial Blood Pressure differs after oxytocin is applied.

Test Statistics<sup>b</sup>

	ABPafter - ABPbefore
Z	-2.934 <sup>a</sup>
Asymp. Sig. (2-tailed)	.003

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

p=0.003 ; Reject  $H_0$

Statistics

	ABPbefore	ABPafter
N	11	11
Valid	0	0
Missing	97.00	75.00
Median		

Conclusion: There is a significant decrease in median blood pressure of pregnant women after oxytocin is applied.

# Lecture 2 of Y1 C3

Chi-square Tests

Chi-square test of independence

Chi-square test of homogeneity

Mc-Nemar Chi-square Test

		<b>Parametric Methods</b>	<b>Non parametric Methods</b>
		<b>Assumption:</b> Data have to be sampled from Gaussian distribution (normally distributed)&populations have equal variances (homogeneity of variances)	No assumption. It is preferred when normality is violated and having small sample size ( $n < 30$ )
Quantitative Data	<b>1 sample vs hypothesized value comparison</b>	One-sample t-test	Signed rank test or Wilcoxon Rank Sum Test
	<b>2 samples Comparisons</b>	Unpaired t-test	Mann-Whitney U Test
		Paired t-test	Wilcoxon Matched Pairs Test
	<b>More than 2 samples Comparisons</b>	Independent Samples  One way-ANOVA Test If $p < 0.05$ according to F-test, Multiple Comparison Tests are needed; such as; Tukey, Bonferroni, Scheffé Multiple Comparison Tests To compare all vs. only with control group, Dunnett Multiple Comparison Test can be used.	Kruskal-Wallis Test If $p < 0.05$ use Dunn Multiple Comparison Test
		Paired Samples or Matched Pair Samples  Repeated Measures ANOVA Test If $p < 0.05$ according to F-test, Multiple Comparison Tests are neededStudent such as Newman-Keuls	Friedman Test If $p < 0.05$ use Dunn Multiple Comparison Test
	Crosstabulation	Independent Samples  ---	<b>X<sup>2</sup>-Test</b> * Homogeneity test *Independence test If the association is significant ; Phi or Cramer's V coefficients can be used In order to measure the degree and to determine the direction of the association.
Categorical (qualitative) Data		Paired Samples or Matched Pair Samples  ---	McNemar X <sup>2</sup> test (2x2) or Stuart-Maxwell test(3x3)
<b>Correlation Coefficient</b> (This measures the degree and determines the direction of linear co-relation between two continuous variables, x and y)		Pearson Correlation Coefficient	Spearman Correlation Coefficient

# Chi-square Tests

			<b>Parametric Methods</b> <i>Assumption: Data have to be sampled from Gaussian distribution (normally distributed) &amp; populations have equal variances (homogeneity of variances)</i>	<b>Non parametric Methods</b> <i>No assumption. It is preferred when normality is violated and having small sample size (n&lt;30)</i>
<b>Categorical (qualitative) Data</b>	<u>Crosstabulation</u>	<i>Independent Samples</i>	---	<b>X<sup>2</sup>-Test</b> * Homogeneity test *Independence test If the association is significant; Phi or Cramer's V coefficients can be used in order to measure the degree and to determine the direction of the association.
		<i>Paired Samples or Matched Pair Samples</i>	---	<b>McNemar X<sup>2</sup> test (2x2) or Stuart-Maxwell test(3x3)</b>

- ▶ Chi-square test of independence / homogeneity
  - Applied for testing the association between two categorical variables for independent samples
  - The same procedure is followed in both tests, but the number of variables & populations, and hypotheses are different.
- ▶ Mc-Nemar Chi-square test
  - Applied for testing the association between two categorical variables for dependent samples

# Purpose of Chi-square Tests

- ▶ A common goal of many research studies is to investigate the association of 2 factors, both being categorical variables.
- ▶ Examples:
  - Is smoking during pregnancy (yes/no) associated with low birth weight (yes/no)?
  - Is income associated with the number of visits to doctor in a year?
    - Income: low/medium/high
    - # of visits: none / 1-2 / 3-4 / >4
- ▶ Analysis of categorical data is necessary
  - Based on frequencies

## Chi-square test of independence

- ▶ we have two characteristics of a population, and we want to see if there is any association between the characteristics
- ▶ (2 variables, 1 population)
- ▶  $H_0$ : the variables are independent
- ▶  $H_1$ : the variables are dependent

## Chi-square test of homogeneity

- ▶ we take samples from different populations, and we want to test to see if the proportions in various categories is the same for each population
- ▶ (1 variable, multiple populations)
- ▶  $H_0$ : populations have the same proportion of individuals with some characteristic.
- ▶  $H_1$ : populations have different proportion of individuals with some characteristic.

# Contingency Tables

- ▶ A contingency table is a cross table showing the distribution of two variables; one in rows and another in columns.
- ▶ Example:

		# of doctor visits				
		None	1-2	3-4	>4	Total
income	Low	54	105	68	33	260
	Medium	40	74	140	30	284
	High	35	85	96	40	256
Total		129	264	304	103	800

# Chi-square testing procedures

- ▶ The idea behind chi-square tests is to compare actual counts to the counts we would expect if the null hypothesis were true.
- ▶ We check if there is a significant difference between the actual counts and expected counts.
- ▶ The method for obtaining the expected counts requires that we determine the number of observations within each cell.

# Expected frequencies in Chi-square tests

- ▶ To find the expected frequencies in a cell when performing a chi-square test, we multiply the row total of the row containing the cell by the column total of the column containing the cell and divide this result by the table total. That is

$$\text{Expected frequency} = \frac{(\text{row total})(\text{column total})}{\text{table total}}$$

# Test statistic for Chi-square tests

- Let  $O_i$  represent the observed number of counts in the  $i^{\text{th}}$  cell,  $E_i$  represent the expected number of counts in the  $i^{\text{th}}$  cell. Then,

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

approximately follows the chi-square distribution with degrees of freedom:

$$\begin{aligned} \text{df} &= (\#\text{rows} - 1)(\#\text{columns} - 1) \\ &= (R-1)(C-1) \end{aligned}$$

# Example

## Chi-square Test of Independence

- In a survey, 883 males and 893 females were asked, ‘If you could have only one of the following, which would you pick: money, health, or love?’
- Their responses are presented in the table below:

	Money	Health	Love	Total
Men	82	446	355	883
Women	46	574	273	893
Total	128	1020	628	1776

- Test the claim that gender and choices are independent at the  $\alpha = 0.05$  level of significance.

# Example

## Chi-square Test of Independence

### Hypotheses:

- ▶  $H_0$ : there is no association between gender and lifestyle choice, the variables are independent
- ▶  $H_1$ : there is an association between gender and lifestyle choice, the variables are dependent

# Example

## Chi-square Test of Independence

- ▶ Calculation of the expected frequencies for each cell in the table

Observed Counts	Money	Health	Love	Total
Men	82	446	355	883
Women	46	574	273	893
Total	128	1020	628	1776

Expected Counts	Money	Health	Love
Men	$\frac{883 * 128}{1776} = 63.64$	$\frac{883 * 1020}{1776} = 507.13$	$\frac{883 * 628}{1776} = 312.23$
Women	$\frac{893 * 128}{1776} = 64.36$	$\frac{893 * 1020}{1776} = 512.87$	$\frac{893 * 628}{1776} = 315.77$

# Example

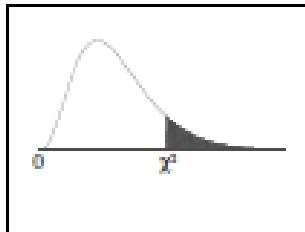
## Chi-square Test of Independence

Computation of the test statistic and P-value:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 36.84$$

- P-value is found by the use of chi-square distribution table with the respective df:  
 $df = (R-1)(C-1) = (2-1)(3-1) = 2$
- P-value < 0.005  $\rightarrow$  how??
  - By using chi-square table

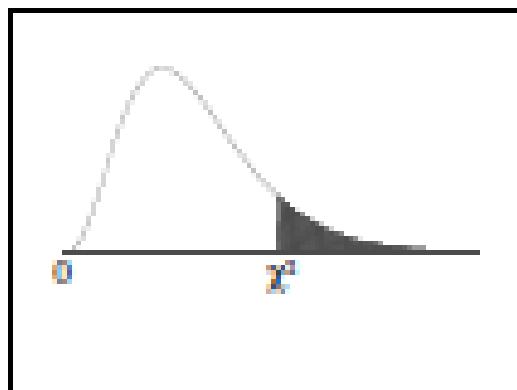
### Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $X^2 = \chi_{\alpha}^2$ .

<i>df</i>	$\chi_{0.00}^2$	$\chi_{0.01}^2$	$\chi_{0.05}^2$	$\chi_{0.10}^2$	$\chi_{0.20}^2$	$\chi_{0.50}^2$	$\chi_{0.90}^2$	$\chi_{0.95}^2$	$\chi_{0.99}^2$	$\chi_{0.995}^2$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.231	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.965
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.309	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.295	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.305	37.166
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.389	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.569
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.193	46.983	49.645
28	12.461	13.563	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.903
70	43.275	45.442	48.758	51.739	55.329	85.627	90.631	95.023	100.426	104.215
80	51.172	53.540	57.153	60.391	64.278	95.678	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

## Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi_{\alpha}^2$ .

$\text{df}$	$\chi_{0.000}^2$	$\chi_{0.001}^2$	$\chi_{0.01}^2$	$\chi_{0.02}^2$	$\chi_{0.05}^2$	$\chi_{0.10}^2$	$\chi_{0.20}^2$	$\chi_{0.50}^2$	$\chi_{0.70}^2$	$\chi_{0.90}^2$	$\chi_{0.95}^2$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	
3	0.072	0.115	0.216	0.362	0.584	6.251	7.815	9.348	11.345	12.838	
4	0.307	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.760	
...	...	...	...	...	...	...	...	...	...	...	...

# Example

## Chi-square Test of Independence

### Statistical Result

- ▶ P-value < 0.005 means p-value <  $\alpha$  and we reject  $H_0$

### Conclusion:

- ▶ With 95% significance, there is evidence that there exists a significant association between gender and lifestyle choice and that these variables are dependent.

# Example

## Chi-square Test of Homogeneity

- ▶ The following question was asked of a random sample of individuals in 1992, 1998, and 2001: “Would you tell me if you feel being a doctor is an occupation of very great prestige?”
- ▶ The results of the survey are presented below:

	1992	1998	2001	Total
Yes	549	539	570	1658
No	522	578	599	1699
Total	1071	1117	1169	3357

- ▶ Test the claim that the proportion of individuals that feel being a doctor is an occupation of very great prestige is the same for each year at the  $\alpha = 0.01$  level of significance.

# Example

## Chi-square Test of Homogeneity

### Hypotheses:

- ▶  $H_0$ : the proportions of individuals who feel being a doctor is an occupation of very great prestige in each year are equal (homogeneous)
- ▶  $H_1$ : the proportions of individuals who feel being a doctor is an occupation of very great prestige in each year are not equal

# Example

## Chi-square Test of Homogeneity

- ▶ Calculation of the expected frequencies for each cell in the table

Observed Counts	1992	1998	2001	Total
Yes	549	539	570	1658
No	522	578	599	1699
Total	1071	1117	1169	3357

Expected Counts	Money	Health	Love
Men	$\frac{1658 * 1071}{3357} = 528.96$	$\frac{1658 * 1117}{3357} = 551.68$	$\frac{1658 * 1169}{3357} = 577.36$
Women	$\frac{1699 * 1071}{3357} = 542.04$	$\frac{1699 * 1117}{3357} = 565.32$	$\frac{1699 * 1169}{3357} = 591.64$

# Example

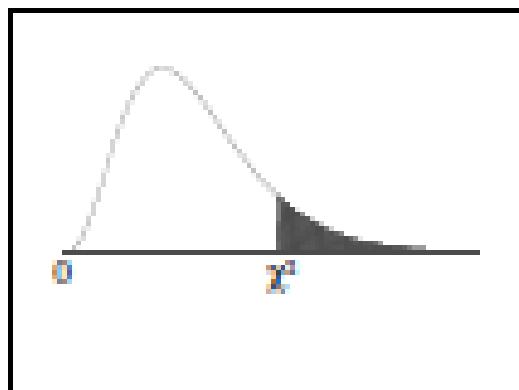
## Chi-square Test of Homogeneity

Computation of the test statistic and P-value:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.26$$

- P-value is found by the use of chi-square distribution table with the respective df:  
 $df = (R-1)(C-1) = (2-1)(3-1) = 2$
- $0.1 < P\text{-value} < 0.9$

## Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi_{\alpha}^2$ .

$df$	$\chi_{0.000}^2$	$\chi_{0.001}^2$	$\chi_{0.01}^2$	$\chi_{0.02}^2$	$\chi_{0.05}^2$	$\chi_{0.10}^2$	$\chi_{0.20}^2$	$\chi_{0.50}^2$	$\chi_{0.70}^2$	$\chi_{0.90}^2$	$\chi_{0.95}^2$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	
3	0.072	0.115	0.216	0.362	0.584	6.251	7.815	9.348	11.345	12.838	
4	0.307	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.760	
...	...	...	...	...	...	...	...	...	...	...	...

# Example

## Chi-square Test of Homogeneity

### Statistical Result

- $0.1 < \text{P-value} < 0.9$  means  $\text{p-value} > \alpha$  and we fail to reject  $H_0$

### Conclusion:

- ▶ With 95% significance, there is NO evidence that the proportions of individuals who feel being a doctor is an occupation of very great prestige is different each year

# Chi-square Tests

			<b>Parametric Methods</b> <b>Assumption:</b> Data have to be sampled from Gaussian distribution (normally distributed) & populations have equal variances (homogeneity of variances)	<b>Non parametric Methods</b> No assumption. It is preferred when normality is violated and having small sample size ( $n<30$ )
<b>Categorical (qualitative) Data</b>	<u>Crosstabulation</u>	<i>Independent Samples</i>	---	<b>X<sup>2</sup>-Test</b> * Homogeneity test *Independence test If the association is significant; Phi or Cramer's V coefficients can be used in order to measure the degree and to determine the direction of the association.
		<i>Paired Samples or Matched Pair Samples</i>	---	McNemar X <sup>2</sup> test (2x2) or Stuart-Maxwell test(3x3)

- ▶ Chi-square test of independence / homogeneity
  - Applied for testing the association between two categorical variables for independent samples
  - The same procedure is followed in both tests, but the number of variables & populations, and hypotheses are different.
- ▶ Mc-Nemar Chi-square test
  - Applied for testing the association between two categorical variables for dependent samples

# Mc-Nemar testing procedures

- ▶ The test statistic of a Mc-Nemar chi-square test is:
- ▶  $\chi^2 = \frac{(b-c)^2}{b+c}$  where b, c are obtained from the contingency table

		Factor 1	
		Yes	No
Factor 2	Yes	a	b
	No	c	d

with degrees of freedom  $df=(R-1)(C-1)$

# Example

## Mc-Nemar Chi-square Test

- A study was conducted on 96 individuals to test their knowledge on the harm of smoking before and after a seminar on this topic, and the results are obtained as:

		After seminar		
		satisfactory	unsatisfactory	
Before seminar	satisfactory	30	25	55
	unsatisfactory	10	31	41
		40	56	96

# Example

## Mc-Nemar Chi-square Test

- ▶ A study was conducted on 96 individuals to test their knowledge on the harm of smoking before and after a seminar on this topic.

### Hypotheses:

- ▶  $H_0$ : there is no difference on the knowledge about the harm of smoking before and after the seminar
- ▶  $H_1$ : there is a difference on the knowledge about the harm of smoking before and after the seminar

# Example

## Mc-Nemar Chi-square Test

		After seminar		
		satisfactory	unsatisfactory	
Before seminar	satisfactory	30	25	55
	unsatisfactory	10	31	41
		40	56	96

►  $\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(25-10)^2}{25+10} = 6.428$

► with  $df=(2-1)(2-1)=1$

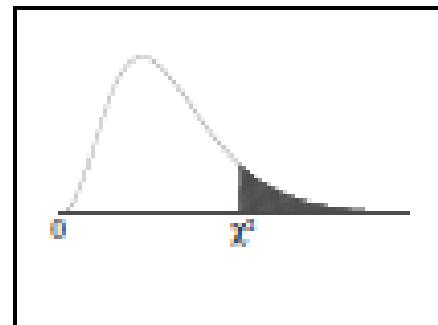
# Example

## Mc-Nemar Chi-square Test

- $\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(25-10)^2}{25+10} = 6.428$
- with  $df=(2-1)(2-1)=1$

Chi-Square Distribution Table

$0.01 < p < 0.025$   
Reject  $H_0$



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi^2_{\alpha}$ .

We conclude that, there is a significant difference on the knowledge about the harm of smoking before and after the seminar

$\alpha$	$\chi^2_{0.05}$	$\chi^2_{0.01}$	$\chi^2_{0.001}$	$\chi^2_{0.0001}$	$\chi^2_{0.05}$	$\chi^2_{0.01}$	$\chi^2_{0.001}$	$\chi^2_{0.0001}$	$\chi^2_{0.05}$	$\chi^2_{0.01}$	$\chi^2_{0.001}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	
2	0.010	0.020	0.031	0.043	0.211	4.207	5.991	7.779	9.210	10.830	
3	0.072	0.115	0.216	0.352	0.584	6.231	7.815	9.348	11.345	12.838	
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	
5	0.412	0.554	0.831	1.146	1.610	9.236	11.070	12.833	15.086	16.760	

# Lecture 3 of Y1 C3

Comparing Risk

Relative Risk (RR)

Odds Ratio (OR)

Some basic concepts for diagnosis tests

# Analyzing 2x2 contingency tables

		Factor 1	
		Yes	No
Factor 2	Yes	a	b
	No	c	d

- ▶ 1) Chi-square tests
- ▶ 2) Estimating Risk
  - Relative Risk
  - Odds Ratio

# Analyzing 2x2 contingency tables

- ▶ There are 3 ways to obtain a 2x2 contingency (cross) table:
  - 1) from a cross-sectional study
  - 2) from a prospective study
    - Cohort studies
  - 3) from a retrospective study
    - Case-control studies

# Recall:

## Cross-sectional studies

- ▶ Prevalence studies
- ▶ Describe a group of subjects at a particular point in time
- ▶ Non-directional studies
- ▶ Example: Association between diabetes and obesity

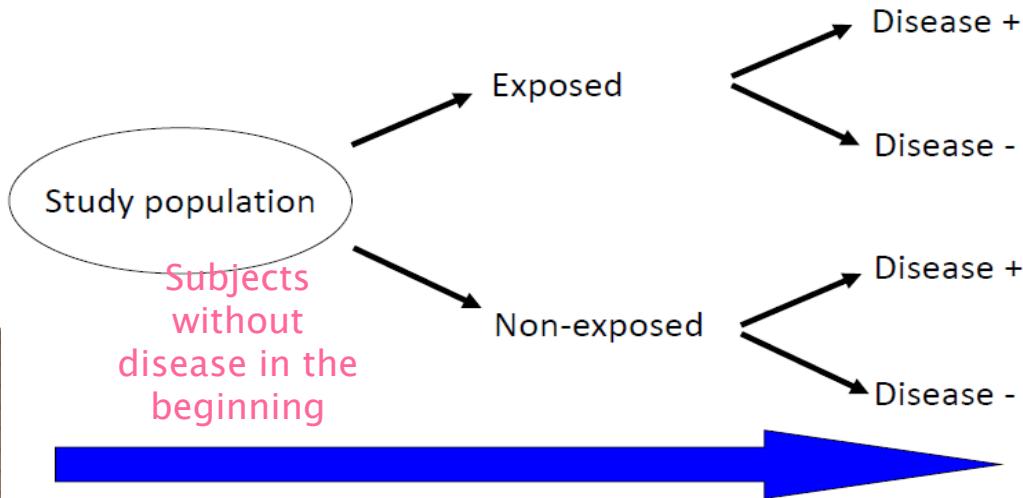
		Obesity	
		Yes	No
Diabetes	Yes	a	b
	No	c	d

Chi-square test of independence is a good way of analyzing such data

# Recall: Cohort studies

- ▶ Find a group of individuals without the disease, and separate them into those
  - With exposure
  - Without exposure
- ▶ Then, follow over time and measure the disease rates in both groups

## Cohort study / Follow-up study



Example:

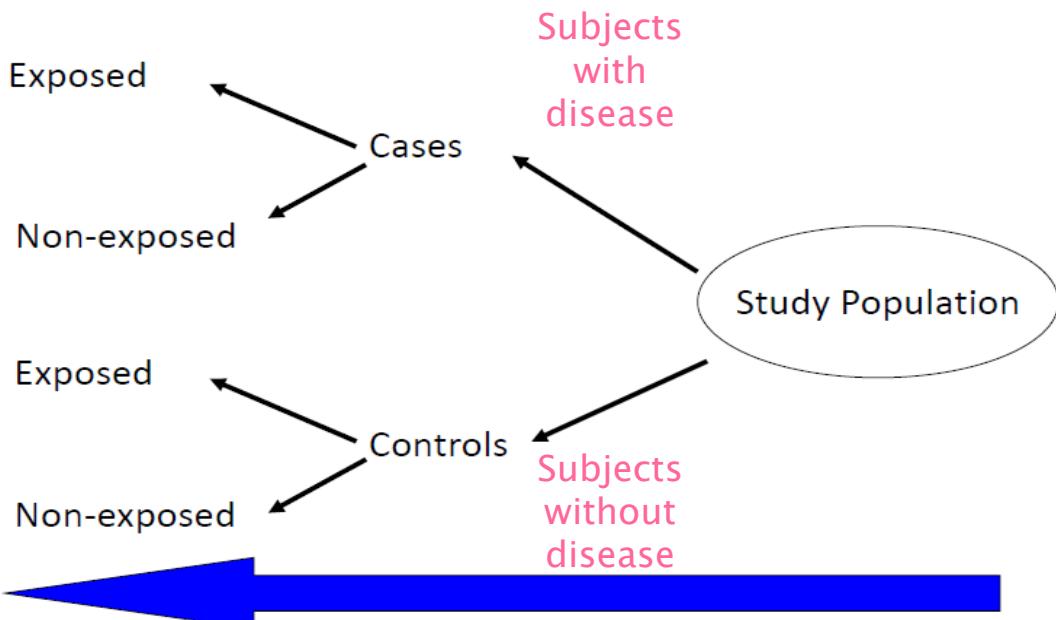
Exposure → smoking  
Disease → lung cancer

		Lung cancer	
		Yes	No
smoking	Yes	a	b
	No	c	d

# Recall: Case-control studies

- ▶ Identify individuals
  - With the disease of interest (case), and
  - Without the disease (control)
- ▶ And then retrospectively find the records of those individuals to fill in the table.

## Case-control study



Example:

Exposure → smoking  
Disease → lung cancer

		Lung cancer	
		Cases (disease +)	Control (disease -)
smoking	Yes	a	b
	No	c	d

# Analyzing 2x2 contingency tables

		Factor 1	
		Yes	No
Factor 2	Yes	a	b
	No	c	d

- ▶ 1) Chi-square tests Cross sectional studies  
Cohort studies  
Case-control studies
- ▶ 2) Estimating Risk
  - Relative Risk Cohort studies
  - Odds Ratio Case-control studies

# Estimating Relative Risk for a Prospective (Cohort) Study

- By calculating the proportions having the outcome (disease +) in each group, we estimate the risk (of having the disease) in one group compared to the other.

		Lung cancer		
		Yes	No	Total
smoking	Yes	30	15	45
	No	20	35	55
	total	50	50	100

$$I_{exposed} = 30/45$$

$$I_{nonexposed} = 20/55$$

$$RR = \frac{I_{exposed}}{I_{nonexposed}} = 1.83$$

What does it mean?

The risk of developing lung cancer is 1.83 times higher in smokers than nonsmokers

# How to further interpret RR?

- ▶ If  $RR < 1$  then exposure is a protective factor
  - ▶ If  $RR = 1$  then there is no effect of the exposure
  - ▶ If  $RR > 1$  then exposure is a risk factor
- 
- ▶ Exposure: smoking
  - ▶ Disease: lung cancer
  - ▶  $RR = 1.83 > 1$ 
    - Smoking is a risk factor of lung cancer

# Finding Odds Ratio for a Retrospective (Case-control) Study

- For case-control studies, we need a method based on calculations within each group.

		Lung cancer		
		Cases (disease +)	Control (disease -)	Total
smoking	Yes	30	15	45
	No	20	35	55
	total	50	50	100

$$Odds_{cases} = 30/20$$

$$Odds_{controls} = 15/35$$

$$OR = \frac{Odds_{cases}}{Odds_{controls}} = \frac{30 \times 35}{20 \times 15} = 3.5$$

What does it mean?

The odds of developing lung cancer is 3.5 times higher in smokers than nonsmokers

# How to further interpret OR?

- ▶ If  $OR < 1$  then exposure is a protective factor
  - ▶ If  $OR = 1$  then there is no effect of the exposure
  - ▶ If  $OR > 1$  then exposure is a risk factor
- 
- ▶ Exposure: smoking
  - ▶ Disease: lung cancer
  - ▶  $OR = 3.5 > 1$ 
    - Smoking is a risk factor of lung cancer

# To sum up..

		disease	
		Yes	No
exposure	Yes	a	b
	No	c	d

- ▶ If it is a cohort study,
- ▶  $RR = \frac{a/a+b}{c/c+d}$
- ▶ If it is a case-control study,
- ▶  $OR = \frac{a/c}{b/d}$
  
- ▶ If  $OR < 1$  then exposure is a protective factor
- ▶ If  $OR = 1$  then there is no effect of the exposure
- ▶ If  $OR > 1$  then exposure is a risk factor

# Some basic concepts for diagnosis tests

- ▶ False Positive
- ▶ False Negative
- ▶ Sensitivity
- ▶ Specificity
- ▶ Predictive Value Positive
- ▶ Predictive Value Negative

# Lecture 4 of Y1 C3

## Correlation Simple Linear Regression

# How do we measure association between two variables?

1. For categorical variables
  - ▶ Relative Risk ( $RR$ )
  - ▶ Odds Ratio ( $OR$ )
2. For continuous variables
  - ▶ Simple Linear Regression
    - Linear model of the association is constructed
  - ▶ Correlation Coefficient  $r$  (or  $R$ )
  - ▶ Coefficient of Determination ( $R^2$ )

# Association between two quantitative variables

- ▶ Simple Linear Regression and Correlation
  - Statistical techniques that use the idea that one variable may be related to one or more variables through an equation.
- ▶ Here we consider the relationship of two variables only in a linear form, which is called linear regression and linear correlation; or **simple regression and correlation**.
- ▶ The relationships between more than two variables are called as multiple regression and correlation.

# Association between two quantitative variables

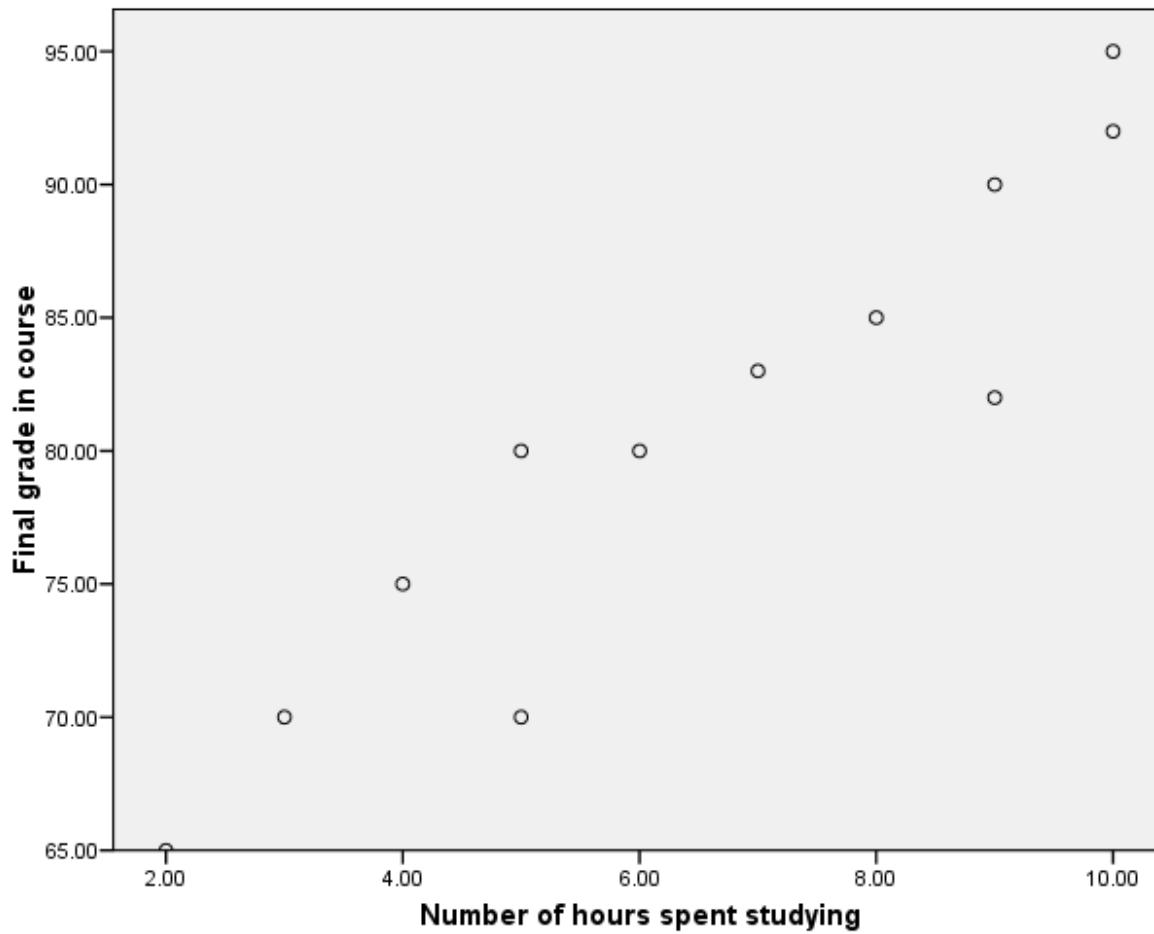
- ▶ Simple regression uses the relationship between the two variables to obtain information about one variable by knowing the values of the other.
- ▶ The equation showing this type of relationship is called *simple linear regression equation*.
- ▶ The related method of correlation is used to measure how strong the relationship is between the two variables is.

# Scatter Plot

- ▶ When we desire to find the relationship between two continuous variables, obtaining the scatter plot of the data helps us to see if there seems to be an association between these variables or not.
- ▶ Scatter diagram plots pairs of bivariate observations  $(x, y)$  on the  $X - Y$  plane
- ▶  $Y$  is called as the dependent variable
- ▶  $X$  is called as an independent variable
- ▶ **The pattern of data is indicative of the type of relationship between two variables:**
  - positive relationship
  - negative relationship
  - no relationship

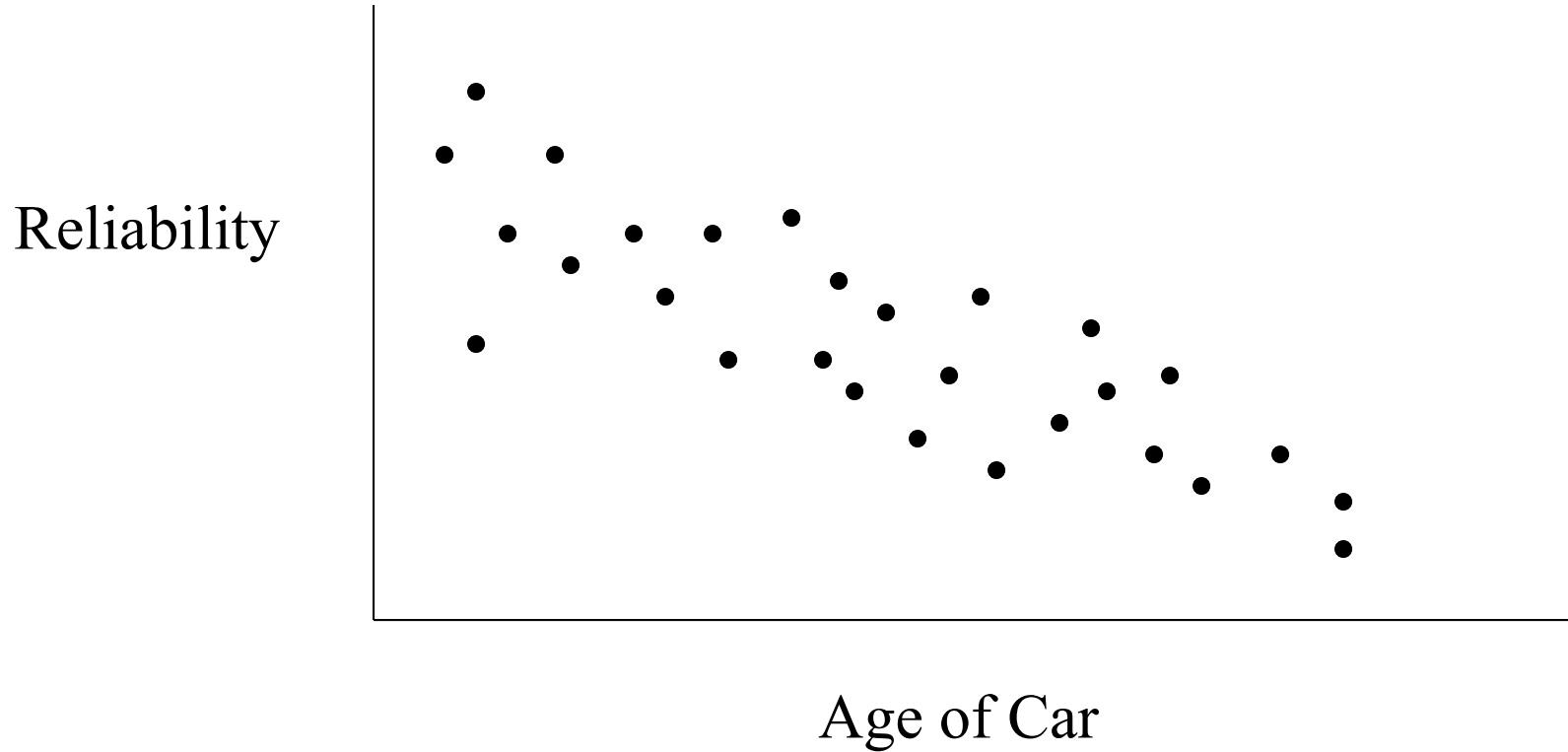
# Scatter Plots

## Positive Relationship



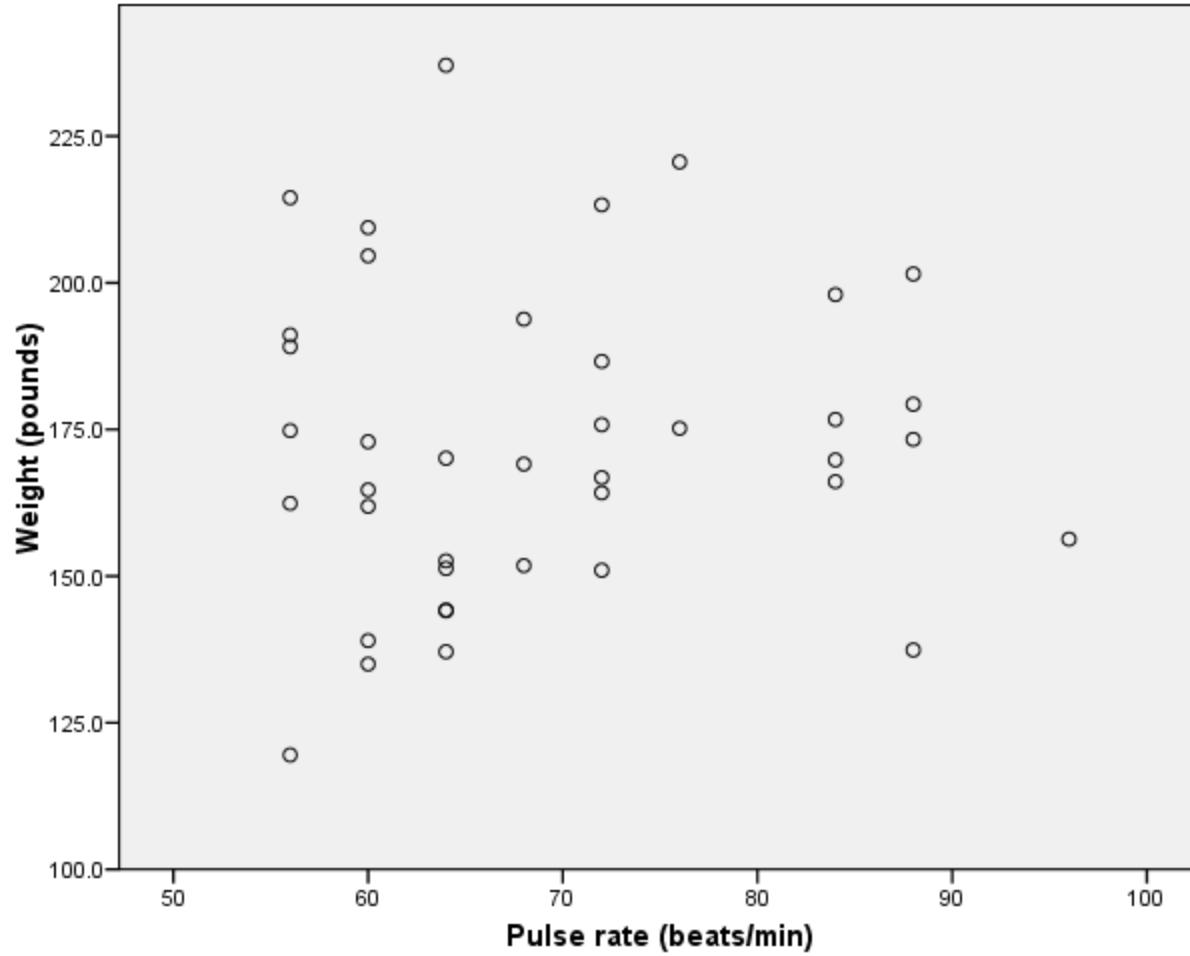
# Scatter Plots

## Negative Relationship



# Scatter Plots

## No Relationship



# Correlation Coefficient ( $r$ )

- ▶  $r$  is a measure that
  - Takes on values between  $-1$  and  $+1$
  - describes and measures two main characteristics of the relationship between two variables X and Y.

## 1) The direction of relationship

- Positive (direct) or negative (indirect)
- $r = 0$  represents no linear relationship between the two variables
- $r > 0$  implies a direct linear relationship
- $r < 0$  implies an inverse(or indirect) linear relationship

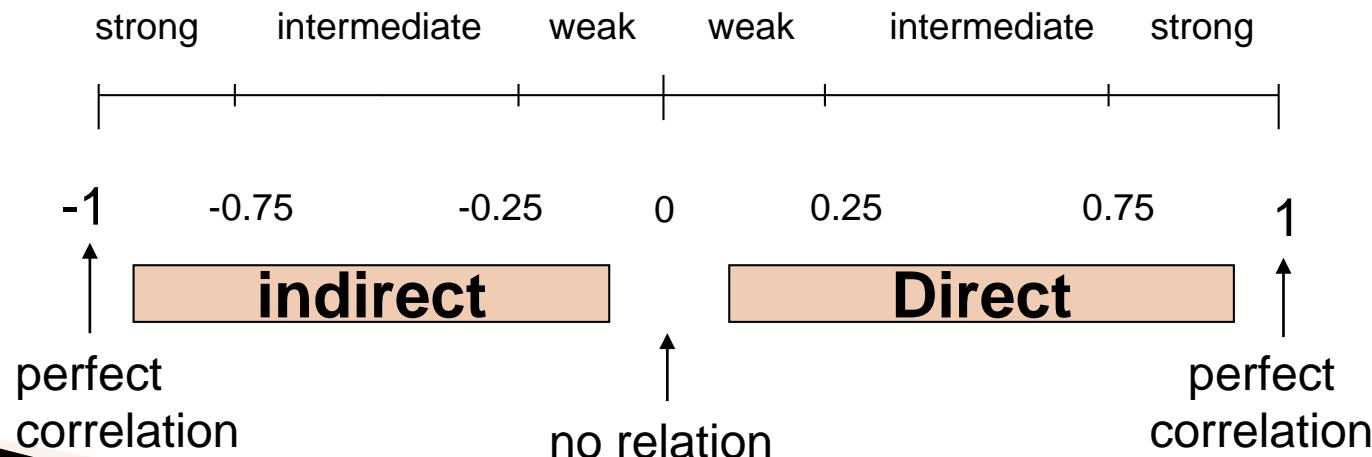
## 2) The strength of the relationship

- The closer  $r$  comes to either  $+1$  or  $-1$ , the stronger is the linear relationship

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

# How to interpret $r$ ?

- ▶  $r = 0$  means no association or correlation between the two variables.
- ▶ If  $0 < r < 0.25 \Rightarrow$  **weak correlation**.
- ▶ If  $0.25 \leq r < 0.75 \Rightarrow$  **intermediate correlation**.
- ▶ If  $0.75 \leq r < 1 \Rightarrow$  **strong correlation**.
- ▶ If  $r = 1 \Rightarrow$  **perfect correlation**.



# Coefficient of Determination ( $R^2$ )

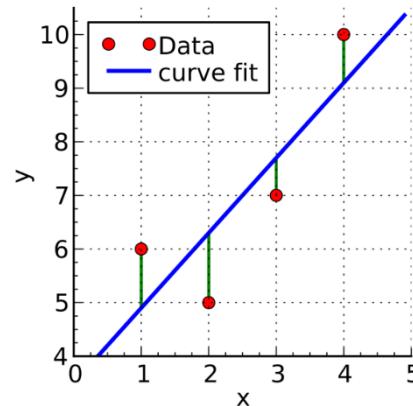
- ▶  $R^2$  is another important measure of linear association between X and Y ( $0 \leq R^2 \leq 1$ )
- ▶ Can be calculated by taking square of  $r$
- ▶  $R^2$  measures the proportion of the total variation in  $Y$  which is explained by  $X$
- ▶  $R^2$  takes on any value between zero and one.
  - $R^2 = 1$ : Perfect match between the line and the data points.
  - $R^2 = 0$ : There are no linear relationship between  $X$  and  $Y$ .

# Simple Linear Regression

- ▶ Suppose that we are interested in a variable  $Y$ , and we want to know about its relationship to another variable  $X$
- OR
- ▶ We want to use  $X$  to predict (or estimate) the value of  $Y$  that might be obtained without actually measuring it, provided the relationship between the two can be expressed by a line.
  
  - ▶ ‘ $X$ ’ is called the **independent variable (predictor)** and
  - ▶ ‘ $Y$ ’ is called the **dependent variable. (outcome)**

# Simple Linear Regression

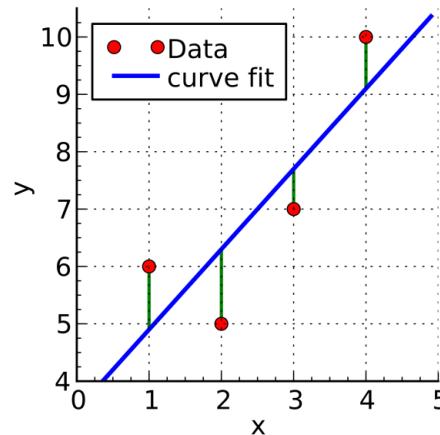
- ▶ An objective, and therefore better, way of determining the position of a straight line is to use the method of least squares.



- ▶ Using this method, we choose a line such that the sum of squares of vertical distances of all points from the line is minimized.
- ▶ These vertical distances, i.e., the distance between y values and their corresponding estimated values on the line are called residuals

# Simple Linear Regression

- ▶ The distance between an actual data point  $Y$  and the predicted point on the line  $\hat{Y}$  is defined as  $Y - \hat{Y}$ . The goal of regression is to find the equation for the line that minimizes these distances.



- ▶ The line which fits the best is called the *regression line* or, sometimes, the *least-squares line*
- ▶ The line always passes through the point defined by the mean of Y and the mean of X

# Simple Linear Regression Procedure

- ▶ The regression model is defined as  $Y = \alpha + \beta X$
- ▶ We select a sample of  $n$  observations  $(x_i, y_i)$  from the population, with the goals
  - Estimate  $\alpha$  and  $\beta$ .
  - Predict the value of  $Y$  at a given value  $x$  of  $X$ .
  - Make tests to draw conclusions about the model and its usefulness.
  - We estimate the parameters  $\alpha$  and  $\beta$  by ' $a$ ' and ' $b$ ' respectively by using sample regression line:
  - $\hat{Y} = a + bX$  where we calculate

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

# Correlation vs. Regression

- ▶ Correlation Coefficient,  $r$  (or  $R$ ), measures the *strength* of bivariate association
- ▶ The regression line is a *prediction equation* that estimates the values of  $y$  for any given  $x$

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$Y = \alpha + \beta X$$

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

# Example

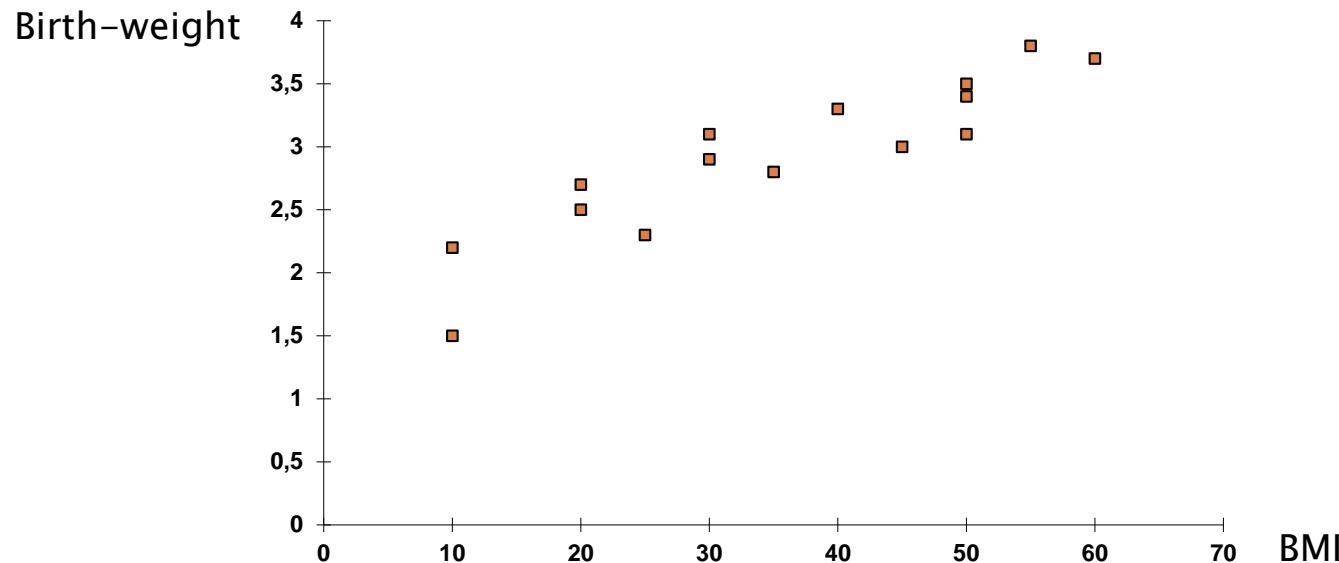
- ▶ A researcher believes that there is a linear relationship between BMI ( $\text{Kg}/\text{m}^2$ ) of pregnant mothers and the birth-weight (BW in Kg) of their newborn
- ▶ The following data set provide information on 15 pregnant mothers who were contacted for this study

BMI ( $\text{Kg}/\text{m}^2$ )	Birth-weight (Kg)
20	2.7
30	2.9
50	3.4
45	3.0
10	2.2
30	3.1
40	3.3
25	2.3
50	3.5
20	2.5
10	1.5
55	3.8
60	3.7
50	3.1
35	2.8

# Example

## Scatter Plot of BMI and Birthweight

- ▶ Is there a linear relationship between BMI and BW?

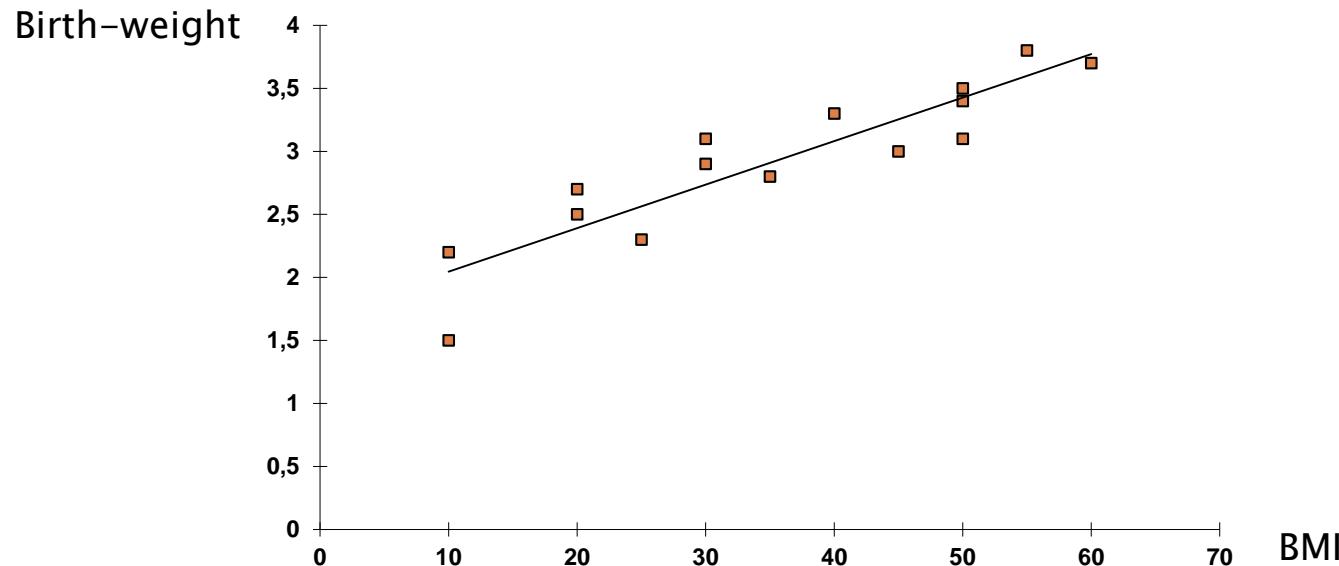


- ▶ Yes, it seems so.
- ▶ Although we could fit a line "by eye" e.g. using a transparent ruler, this would be a subjective approach and therefore unsatisfactory.

# Example

## Scatter Plot of BMI and Birthweight

- ▶ Is there a linear relationship between BMI and BW?



- ▶ Yes, it seems so.
- ▶ Although we could fit a line "by eye" e.g. using a transparent ruler, this would be a subjective approach and therefore unsatisfactory.

# Example

## Simple Linear Regression and Correlation Analysis

- Calculations from the table:

- $\hat{Y} = a + bX$  where

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$b = 0.0345$$

$$a = 1.6994$$

$$R = 0.9074$$

$$R^2 = 0.8234$$

BMI (Kg/m <sup>2</sup> )	Birth-weight (Kg)	x-xbar	y-ybar	(x-xbar)(y-ybar)	(x-xbar) <sup>2</sup>	(y-ybar) <sup>2</sup>
20	2.7	-15.33	-0.22	3.37	235.11	0.0484
30	2.9	-5.33	-0.02	0.11	28.44	0.0004
50	3.4	14.67	0.48	7.04	215.11	0.2304
45	3	9.67	0.08	0.77	93.44	0.0064
10	2.2	-25.33	-0.72	18.24	641.78	0.5184
30	3.1	-5.33	0.18	-0.96	28.44	0.0324
40	3.3	4.67	0.38	1.77	21.78	0.1444
25	2.3	-10.33	-0.62	6.41	106.78	0.3844
50	3.5	14.67	0.58	8.51	215.11	0.3364
20	2.5	-15.33	-0.42	6.44	235.11	0.1764
10	1.5	-25.33	-1.42	35.97	641.78	2.0164
55	3.8	19.67	0.88	17.31	386.78	0.7744
60	3.7	24.67	0.78	19.24	608.44	0.6084
50	3.1	14.67	0.18	2.64	215.11	0.0324
35	2.8	-0.33	-0.12	0.04	0.11	0.0144
35.333333	2.92			126.90	3673.33	5.32
xbar	ybar			SP	SSx	SSy

Thus, linear regression model is:

$$\hat{Y} = 1.6994 + 0.0345X$$

# Example

## Inferences from the model

- From the regression model, we are able to make predictions for some  $X$  values to estimate  $Y$ s
- $\hat{Y} = 1.6994 + 0.0345X$
- This equation allows us to estimate BW of other newborns when the BMI is given.
- For example, for a mother who has BMI=40, i.e.  $X = 40$  we predict BW to be
- $\hat{Y} = 1.6994 + 0.0345X = 1.6994 + 0.0345(40) = 3.0794$

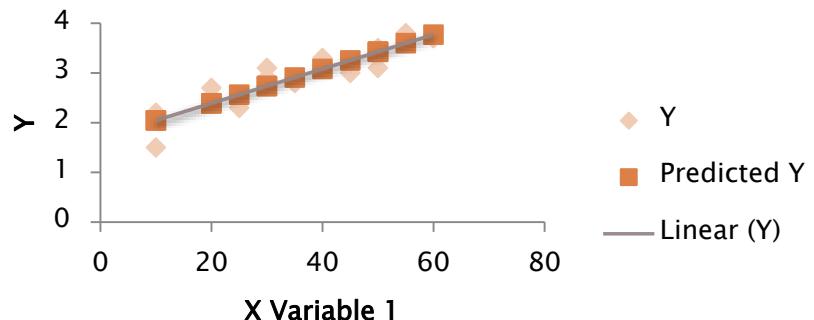
# Example

## Inferences from R and $R^2$

- $R$  – *The correlation coefficient*
  - $R = 0.9074$  indicates a direct linear relationship between BMI and BW
- $R^2$  – *The coefficient of determination*
  - $R^2 = 0.8234$  indicates that 82.34% of the variation in BW is explained by the independent variable  $X$  (BMI).

# Example Excel output for the same data

X Variable 1 Line Fit Plot



SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0,907429					
R Square	0,823427					
Adjusted R Sq	0,809844					
Standard Error	0,268912					
Observations	15					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	4,383923	4,383923	60,62375	3,00469E-06	
Residual	13	0,940077	0,072314			
Total	14	5,324				
	Coefficients	standard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1,699365	0,171458	9,911248	2E-07	1,328951855	2,069778
X Variable 1	0,034546	0,004437	7,786125	3E-06	0,024960934	0,044132
					0,024961	0,044132

# Lecture 5 of Y1 C3

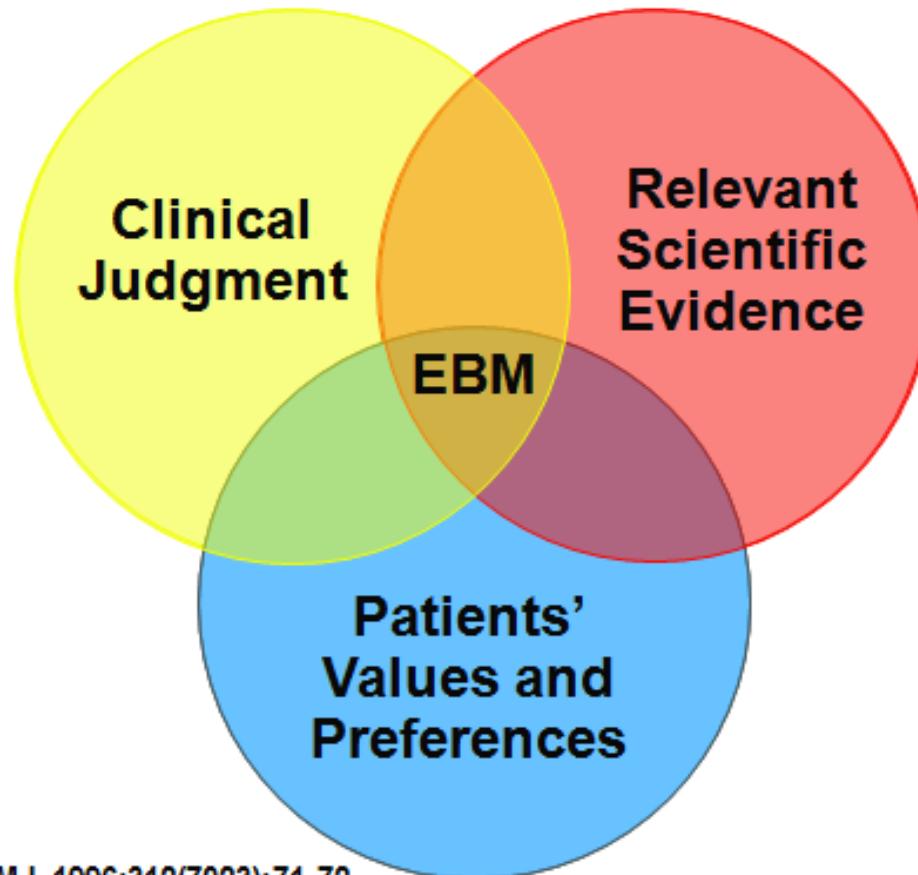
Evidence-Based Medicine  
Systematic Reviews and Meta Analysis

# Evidence-Based Medicine

# Evidence-Based Medicine

- ▶ What is Evidence-Based Medicine(EBM)?
  - Sackett et al. (1996) described EBM as:
    - The conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients
- ▶ «Evidence-based medicine is the integration of best research evidence with clinical expertise and patient values»

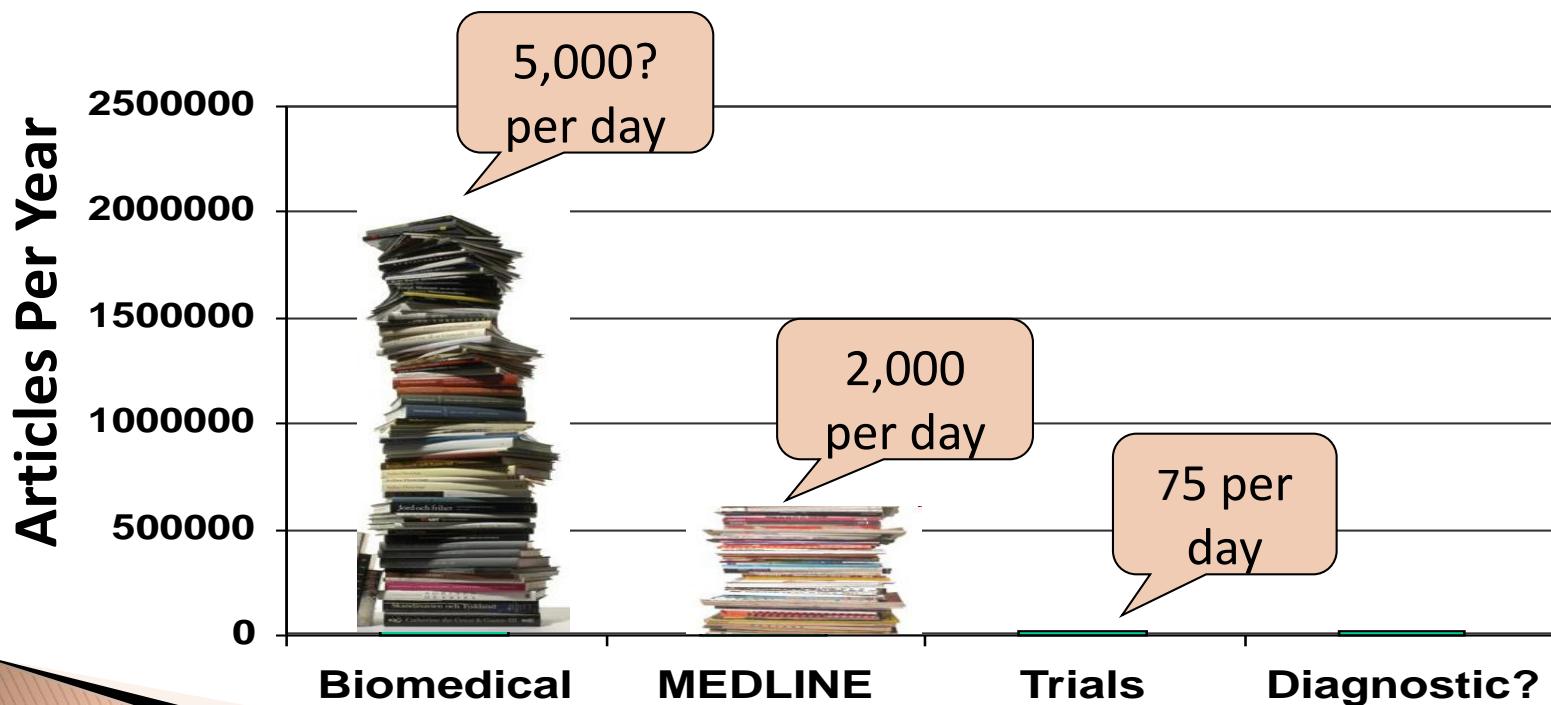
# Evidence-Based Medicine



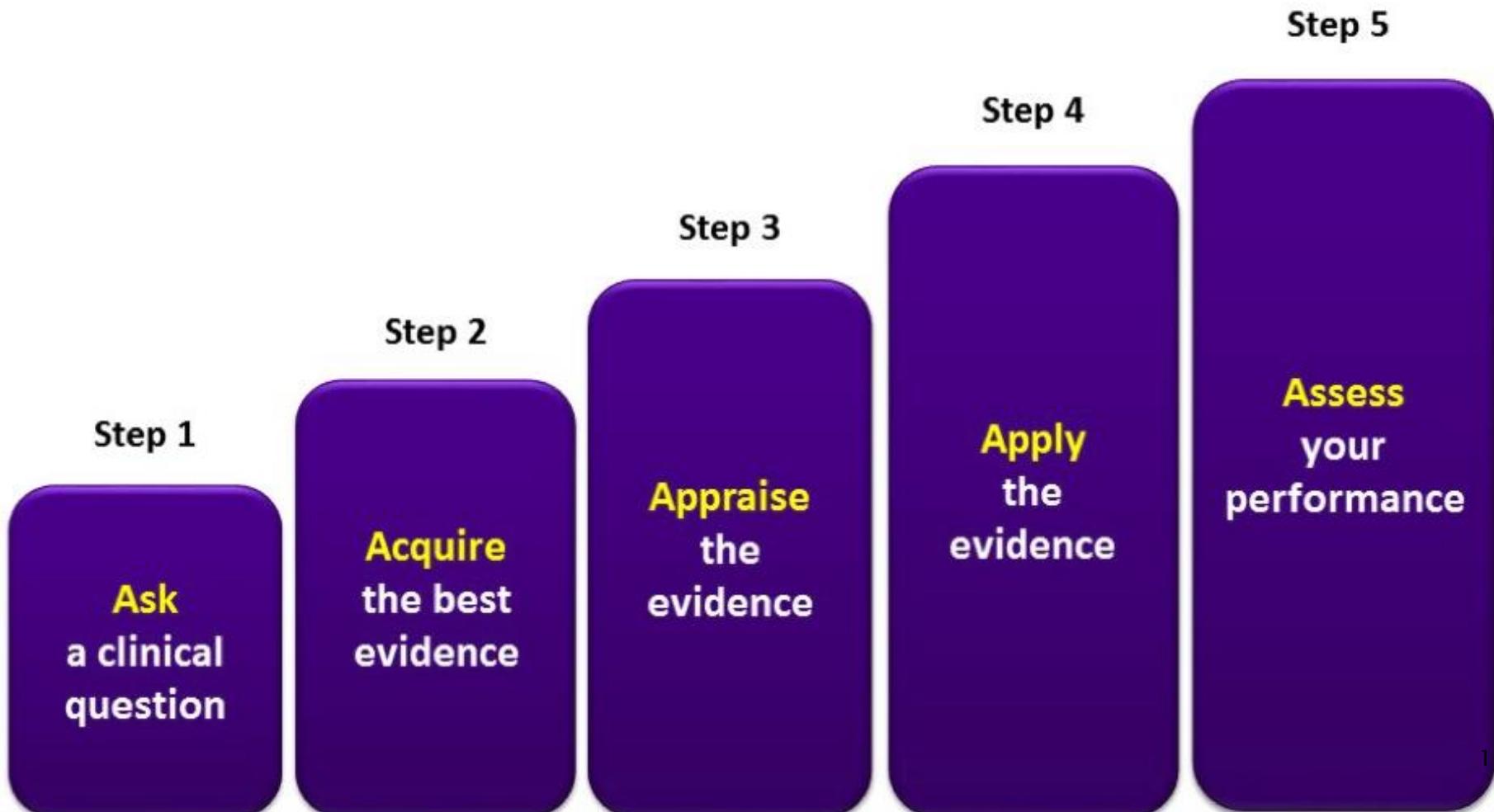
Sackett DL, et al. BMJ. 1996;312(7023):71-72.

# Why do we need to use evidence efficiently?

- ▶ Because there is an epidemic of evidence we need to keep up with and we cannot do this without new skills



# Practicing EBM in 5 steps



# Evidence-Based Medicine

[Evidence-Based Medicine Home](#)

[Steps to Evidence-Based Medicine](#)

[Diagnosis and Therapy Cases](#)

[Online Calculators](#)

[Library Assignment](#)

[Therapy Worksheet](#)

[Diagnosis Worksheet](#)

## Steps to Evidence-Based Medicine

The [MD Program](#) curriculum at the University of Wisconsin School of Medicine and Public Health in Madison includes training in evidence-based medicine (EBM). The following are steps to evidence-based medicine.

### Step 1: Define a Clinical Question

[Formulating Answerable Clinical Questions](#) (Centre for Evidence-Based Medicine)

### Step 2: Find the Evidence

Use the terms from your PICO question to search a current literature database. When using PubMed, the easiest approach is to insert the most important words from your question with and between each one.

Use the [Clinical Queries section of PubMed](#) to search for evidence.

If you are not familiar with searching in PubMed, work through the [National Library of Medicine's online tutorial](#) or contact the medical education liaison librarian at [esevetson@library.wisc.edu](mailto:esevetson@library.wisc.edu).

### If Your Search Yields Too Much Information, Try These Approaches

- Make sure you are using the "specificity" button with either the "diagnosis" or "therapy" button
- Add in further search terms with other words of your PICO question
- Use the limits available when searching

### If Your Search Yields Too Little Information, Try These Approaches

- Use the "sensitivity" button with either the "diagnosis" or "therapy" button
- Remove some of the search terms from your entry
- Remove limits from your search

### Advice on Choosing an Article

Sometimes more than one article fulfills your criteria. With quick review of the abstract you can usually eliminate some of the articles. We suggest you choose an article that:

- Is recent
- Indicates the study is conducted in a blinded fashion
- Clearly compares your diagnostic test or therapy in question to a reference standard
- Comes from a reputable journal

### Step 3: Assess the evidence

[Critical Appraisal of the Evidence](#): After finding either diagnosis or therapy evidence, decide whether it is both valid, or close to the truth and important, or will make a difference in patient care. The order in which validity and importance are considered depends on individual preference, but both should be addressed before deciding whether the evidence can be applied to individual patients.

### Step 4: Apply the Evidence

[Applying Evidence to Patients](#) (Centre for Evidence-Based Medicine)

### Step 5: Communication

Finding and evaluating clinical information is only part of EBM. It is equally important to be able to communicate this information to patients and to integrate patient values and preferences in decision-making<sup>1</sup>. These skills will become more important as external forces, such as managed care, direct-to-consumer advertising and Internet-accessible patient information, become greater factors in doctor-patient interactions. In Patient, Doctor and Society 4, you will be learning and practicing these important skills.

1. McAlister FA, Straus SE, Guyatt GH, Haynes RB for the Evidence-Based Working Group. User's guide to the medical literature: XX. Integrating research evidence with the care of the individual patient. JAMA. 2000;283(1):2829-36.

# EBM - Step1: Ask a Clinical Qn

- ▶ Types of questions:
  - How common is the problem?
    - Prevalence
  - Is early detection worthwhile?
    - Screening
  - Is the diagnostic test accurate?
    - Diagnosis
  - What will happen if we do nothing?
    - Prognosis
  - Does this intervention help?
    - Treatment
  - What are the harms of an intervention?
    - Harms

# EBM – Step1: Ask a Clinical Qn

## ► Formulate the problem by using PICO

P

- **Patient, Population or Problem**
- How would you describe a group of patients similar to yours? What are the most important characteristics of the patient?

I

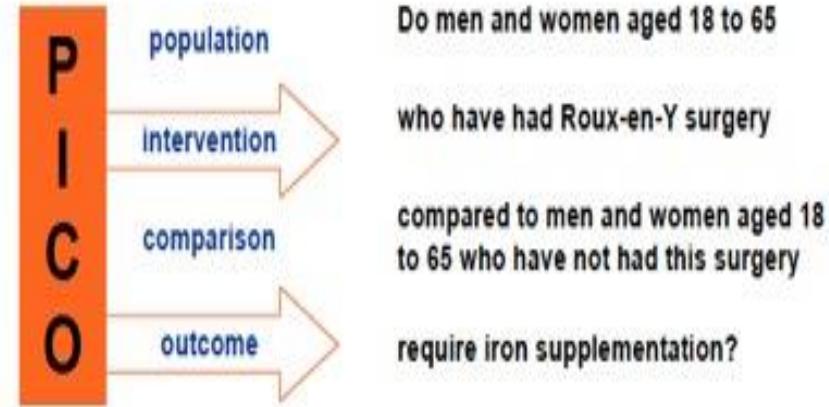
- **Intervention, prognostic factor, or exposure**
- Which main intervention, prognostic factor, or exposure are you considering? What do you want to do for the patient?

C

- **Comparison**
- What is the main alternative to compare with the intervention?

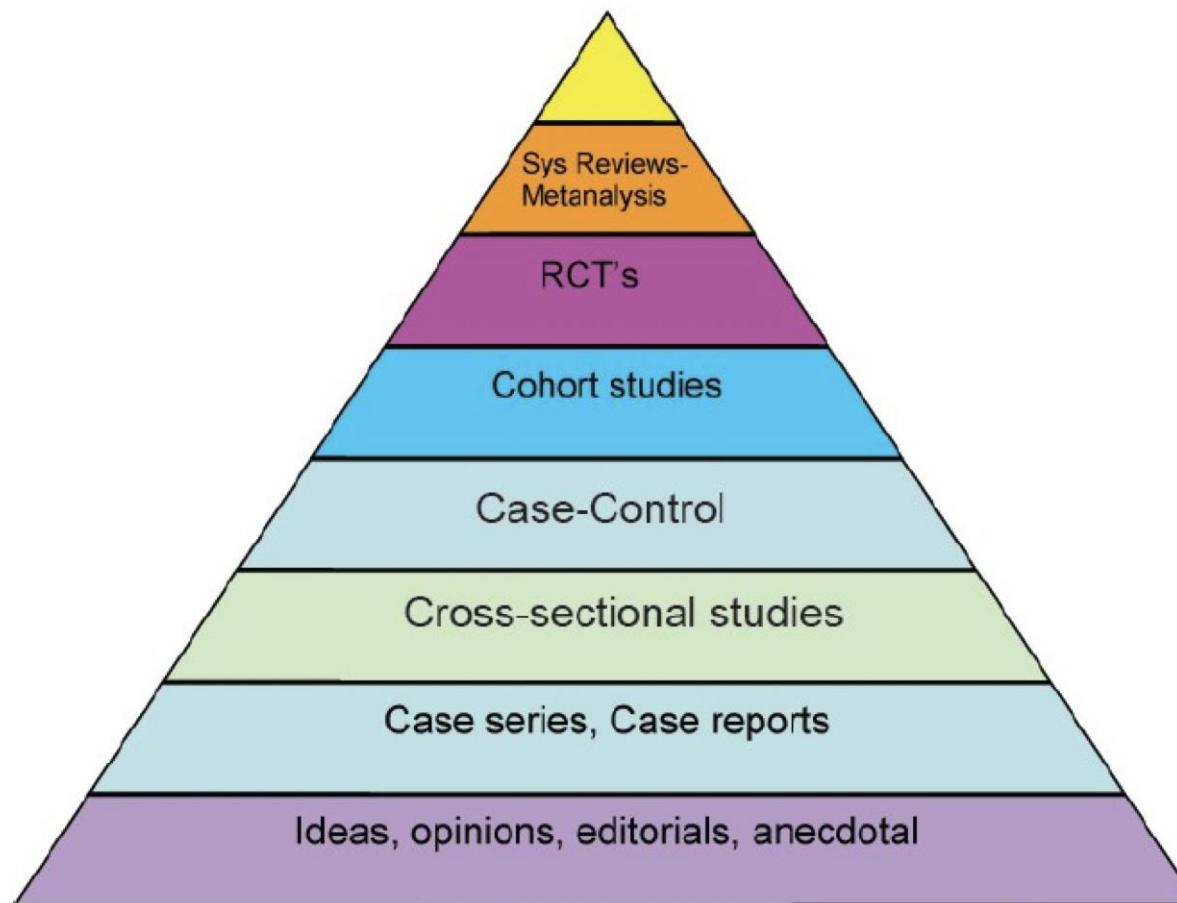
O

- **Outcome**
- What can you hope to accomplish, measure, improve or affect? What are you trying to do for the patient?

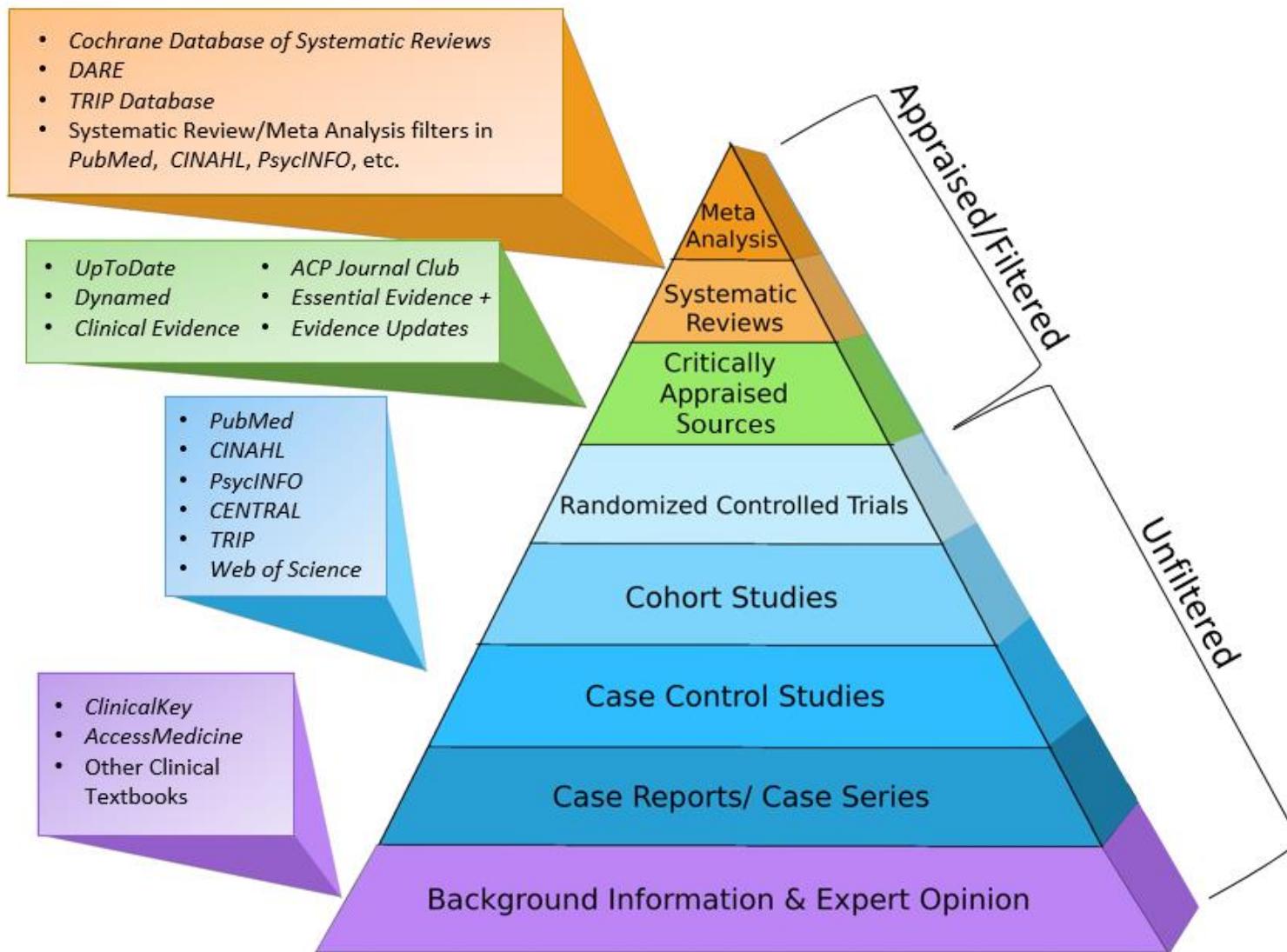


# EBM – Step2: Acquire the best evidence

- ▶ By considering the evidence pyramid



# Evidence Pyramid



# EBM – Step3: Appraise the evidence

- ▶ Critically appraise the methods in order to assess the validity (closeness to the truth) of the evidence (by first checking its abstract)
- ▶ The following questions should be asked:
  - Have all important *outcomes* been considered?
  - Was the study conducted using an appropriate spectrum of patients?
  - Do the results make biological sense?
  - Was the study designated to eliminate bias?

# EBM – Step4: Apply the evidence

- ▶ If you think the results in the source(evidence) that you have found is going to help in caring for your patients, you must ensure that:
  - your patient is similar to those on whom the results were obtained
  - the results can be applied to your patient
  - all clinically important outcomes have been considered
  - the likely benefits are worth the potential harms and costs.

# EBM – Step5: Assess your performance

- ▶ Evaluate your own performance
- ▶ Self-evaluation involves questioning your abilities to complete tasks 1 to 5 successfully.
- ▶ Are you then able to integrate the critical appraisal into clinical practice?
- ▶ You should also ask yourself whether you have learnt from past experience so that the whole process of EBM is now more efficient easier.

# Systematic Reviews and Meta Analysis

# Systematic Reviews

- ▶ The assembly, critical appraisal, and synthesis of all relevant studies that address a specific question.
- ▶ Scientific strategies are applied in ways that limit bias.

# The need for Systematic Reviews

- ▶ There is a large amount of information including evidence from healthcare research which is hard to manage.
- ▶ A clinician/health professional might not have the time, skills or resources to synthesize studies in the literature.
- ▶ Systematic reviews respond to this challenge by identifying, evaluating and synthesizing research-based evidence and present it in an accessible format

# Systematic Reviews

- ▶ Cochrane Database of Systematic Reviews
  - The Cochrane Database of Systematic Reviews (CDSR) is the leading resource for systematic reviews in health care.



Trusted evidence.  
Informed decisions.  
Better health.

[Cochrane.org](#)

Search title, abstract, keyword 

[Browse](#) [Advanced Search](#)

Cochrane Reviews ▾

Trials ▾

More Resources ▾

About ▾

Help ▾

 [Preview the new Cochrane Library website](#)

## Cochrane Database of Systematic Reviews: all issues

Browse issues

2018, Issue 4

[Reviews](#) | [Withdrawn Reviews](#) | [Protocols](#) | [Withdrawn Protocols](#)

- 2018
  - ↳ ● Issue 4 - [Current issue](#)
  - Issue 3
  - Issue 2
  - Issue 1

Since Issue 6, 2013, Cochrane Reviews and Protocols have been published continuously and included in monthly issues. CDSR was published monthly from 2010 and quarterly prior to that. For information on pre-2003 issues, please contact the Cochrane Editorial Unit  
[ceu@cochrane.org](mailto:ceu@cochrane.org)

# Systematic Reviews

## Example 1

- ▶ Does the regular wearing of ultraviolet-blocking sunscreen prevent melanoma?
  - An exhaustive literature search was conducted, resulting in 54 studies on sunscreen and melanoma.
  - Each study was then evaluated to determine whether the study focused specifically on ultraviolet-blocking sunscreen and melanoma prevention; 30 of the 54 studies were retained.
  - The thirty studies were reviewed and showed a strong positive relationship between daily wearing of sunscreen and a reduced diagnosis of melanoma.

# Systematic Reviews

## Example 2

- ▶ Pittler MH, Guo R, Ernst E. Hawthorn extract for treating chronic heart failure. Cochrane Database of Systematic Reviews 2008, Issue 1. Art. No.: CD005312.
- ▶ This systematic review analyzed fourteen studies (randomized, double-blinded, and placebo controlled) that used hawthorn leaf and flower extract monopreparations to determine whether there is any benefit or harm in using hawthorn extract to treat chronic heart failure when compared to placebo. The authors determined that hawthorn extract provides “significant benefit in symptom control and physiologic outcomes” when used as an adjuvant treatment for chronic heart failure.

# Systematic Reviews

## Example 2

Cochrane Database Syst Rev. 2008 Jan 23;(1):CD005312. doi: 10.1002/14651858.CD005312.pub2.

### Hawthorn extract for treating chronic heart failure.

Pittler MH, Guo R, Ernst E.

#### Abstract

**BACKGROUND:** Hawthorn extract is advocated as an oral treatment option for chronic heart failure. Also, the German Commission E approved the use of extracts of hawthorn leaf with flower in patients suffering from heart failure graded stage II according to the New York Heart Association.

**OBJECTIVES:** To assess the benefits and harms as reported in double-blind randomised clinical trials of hawthorn extract compared with placebo for treating patients with chronic heart failure.

**SEARCH STRATEGY:** We searched CENTRAL on The Cochrane Library (issue 2, 2006), MEDLINE (1951 to June 2006), EMBASE (1974 to June 2006), CINAHL (1982 to June 2006) and AMED (1985 to June 2006). Experts and manufacturers were contacted. Language restrictions were not imposed.

**SELECTION CRITERIA:** To be included, studies were required to state that they were randomised, double-blind, and placebo controlled, and used hawthorn leaf and flower extract monopreparations.

**DATA COLLECTION AND ANALYSIS:** Two reviewers independently performed the selection of studies, data extraction, and assessment of methodological quality. Data were entered into RevMan 4.2 software. Results from continuous data were reported as weighted mean difference (WMD) with 95% confidence interval (CI). Where data were suitable for combining, pooled results were calculated.

**MAIN RESULTS:** Fourteen trials met all inclusion criteria and were included in this review. In most of the studies, hawthorn was used as an adjunct to conventional treatment. Ten trials including 855 patients with chronic heart failure (New York Heart Association classes I to III) provided data that were suitable for meta-analysis. For the physiologic outcome of maximal workload, treatment with hawthorn extract was more beneficial than placebo (WMD (Watt) 5.35, 95% CI 0.71 to 10.00, P < 0.02, n = 380). Exercise tolerance were significantly increased by hawthorn extract (WMD (Watt x min) 122.76, 95% CI 32.74 to 212.78, n = 98). The pressure-heart rate product, an index of cardiac oxygen consumption, also showed a beneficial decrease with hawthorn treatment (WMD (mmHg/min) -19.22, 95% CI -30.46 to -7.98, n = 264). Symptoms such as shortness of breath and fatigue improved significantly with hawthorn treatment as compared with placebo (WMD -5.47, 95% CI -8.68 to -2.26, n = 239). No data on relevant mortality and morbidity such as cardiac events were reported, apart from one trial, which reported deaths (three in active, one in control) without providing further details. Reported adverse events were infrequent, mild, and transient; they included nausea, dizziness, and cardiac and gastrointestinal complaints.

**AUTHORS' CONCLUSIONS:** These results suggest that there is a significant benefit in symptom control and physiologic outcomes from hawthorn extract as an adjunctive treatment for chronic heart failure.

# Systematic Reviews: Benefits

- ▶ They are considered at the top of the evidence pyramid (very high level of evidence)
- ▶ They have influence on decision makers
- ▶ They have influence on the design of future studies
- ▶ They cost very little money than a new RCT or observational study
- ▶ They cost less time than a new RCT or observational study

# Systematic Reviews: Benefits

- ▶ Small but clinically significant treatment effects can be detected.
- ▶ The boundaries of what is known and not known are specified.
- ▶ Systematic reviews can help clinicians/scientists by:
  - summarizing large bodies of evidence
  - helping to explain differences among studies on the same question

# Types of Systematic Reviews: Qualitative VS Quantitative

## Qualitative systematic review

- The results of primary studies are summarized
- Not statistically combined
- Described narratively
- Still use other methods to limit bias

## Quantitative systematic review (meta-analysis)

- The results of two or more primary studies are combined
- Statistically combined
  - ▣ Individual patient data (IPD) / Pooled analysis
  - ▣ Aggregate patient data (APD)
  - ▣ Use methods to limit bias

# Meta Analysis Example



## NIH Public Access Author Manuscript

*Nat Genet.* Author manuscript; available in PMC 2010 March 11.

Published in final edited form as:

*Nat Genet.* 2008 December ; 40(12): 1426–1435. doi:10.1038/ng.262.

## Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer

COGENT Study<sup>1</sup>

### Abstract

Genome-wide association (GWA) studies have identified multiple loci at which common variants modestly influence the risk of developing colorectal cancer (CRC). To enhance power to identify additional loci with similar effect sizes, we conducted a meta-analysis of two GWA studies, comprising 13,315 individuals genotyped for 38,710 common tagging SNPs. We undertook replication testing in up to eight independent case-control series comprising 27,418 subjects. We identified four previously unreported CRC risk loci at 14q22.2 (rs4444235, *BMP4*;  $P = 8.1 \times 10^{-10}$ ), 16q22.1 (rs9929218, *CDH1*;  $P = 1.2 \times 10^{-8}$ ), 19q13.1 (rs10411210, *RHPN2*;  $P = 4.6 \times 10^{-9}$ ) and 20p12.3 (rs961253;  $P = 2.0 \times 10^{-10}$ ). These findings underscore the value of large sample series for discovery and follow-up of genetic variants contributing to the etiology of CRC.

# Meta Analysis Example

## Does Problem-based Learning Work? A Meta-analysis of Evaluative Research

DAVID T. A. VERNON, PhD, and ROBERT L. BLAKE, MD

**Abstract**—The purpose of this review is to synthesize all available evaluative research from 1970 through 1992 that compares problem-based learning (PBL) with more traditional methods of medical education. Five separate meta-analyses were performed on 35 studies representing 19 institutions. For 22 of the studies (representing 14 institutions), both effect-size and supplementary vote-count analyses could be performed; otherwise, only supplementary analyses were performed. PBL was found to be significantly superior with respect to students' program evaluations (i.e., students' attitudes and opinions about their programs)— $dw$  (standardized differences between means, weighted by sample size) = +.55,  $CI_{.95}$  = +.40 to +.70—and measures of students' clinical performance ( $dw$  = +.28,  $CI_{.95}$  = +.16 to +.40). PBL and traditional methods did not differ on miscellaneous tests of factual knowledge ( $dw$  = -.09,  $CI_{.95}$  =

+.06 to -.24) and tests of clinical knowledge ( $dw$  = +.08,  $CI_{.95}$  = -.05 to +.21). Traditional students performed significantly better than their PBL counterparts on the National Board of Medical Examiners Part I examination—NBME I ( $dw$  = -.18,  $CI_{.95}$  = -.10 to -.26). However, the NBME I data displayed significant overall heterogeneity ( $Q_t$  = 192.23,  $p < .001$ ) and significant differences among programs ( $Q_b$  = 59.09,  $p < .001$ ), which casts doubt on the generality of the findings across programs. The comparative value of PBL is also supported by data on outcomes that have been studied less frequently, i.e., faculty attitudes, student mood, class attendance, academic process variables, and measures of humanism. In conclusion, the results generally support the superiority of the PBL approach over more traditional methods. *Acad. Med.* 68 (1993):550–563.

# Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials

H A Bischoff-Ferrari, director of centre on aging and mobility,<sup>1,2</sup> B Dawson-Hughes, director of bone metabolism laboratory,<sup>3</sup> H B Staehelin, professor emeritus,<sup>4</sup> J E Orav, associate professor of biostatistics,<sup>5</sup> A E Stuck, professor of geriatrics,<sup>6</sup> R Theiler, head of rheumatology,<sup>7</sup> J B Wong, professor of medicine,<sup>8</sup> A Egli, fellow,<sup>1</sup> D P Kiel, associate professor of medicine,<sup>9</sup> J Henschkowsky, fellow<sup>1,6</sup>

<sup>1</sup>Centre on Aging and Mobility, University of Zurich, Switzerland

<sup>2</sup>Department of Rheumatology and Institute of Physical Medicine, University Hospital Zurich, Switzerland

<sup>3</sup>Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA

<sup>4</sup>Department of Geriatrics, University Hospital Basel, Switzerland

<sup>5</sup>Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

<sup>6</sup>Department of Geriatrics, Inselspital and Spital Netz Bern Zieger, University of Bern, Switzerland

<sup>7</sup>Department of Rheumatology, Triemli Hospital, Zurich, Switzerland

<sup>8</sup>Division of Clinical Decision Making, Tufts Medical Center, Boston, MA, USA

<sup>9</sup>Hebrew Senior Life Institute for Aging Research, Harvard Medical School, Boston, MA, USA

Correspondence to: H A Bischoff-Ferrari, University Hospital Zurich, Gloriastrasse 25, 8091 Zurich, Switzerland  
heike.bischoff@usz.ch

Cite this as: BMJ 2009;339:b3692  
doi:10.1136/bmj.b3692

## ABSTRACT

**Objective** To test the efficacy of supplemental vitamin D and active forms of vitamin D with or without calcium in preventing falls among older individuals.

**Data sources** We searched Medline, the Cochrane central register of controlled trials, BIOSIS, and Embase up to August 2008 for relevant articles. Further studies were identified by consulting clinical experts, bibliographies, and abstracts. We contacted authors for additional data when necessary.

**Review methods** Only double blind randomised controlled trials of older individuals (mean age 65 years or older) receiving a defined oral dose of supplemental vitamin D (vitamin D<sub>3</sub> (cholecalciferol) or vitamin D<sub>2</sub> (ergocalciferol)) or an active form of vitamin D (1α-hydroxyvitamin D<sub>3</sub> (1α-hydroxycalciferol) or 1,25-dihydroxyvitamin D<sub>3</sub> (1,25-dihydroxycholecalciferol)) and with sufficiently specified fall assessment were considered for inclusion.

**Results** Eight randomised controlled trials (n=2426) of supplemental vitamin D met our inclusion criteria.

Heterogeneity among trials was observed for dose of vitamin D (700-1000 IU/day v 200-600 IU/day; P=0.02) and achieved 25-hydroxyvitamin D<sub>3</sub> concentration (25(OH)D concentration: <60 nmol/l v ≥60 nmol/l; P=0.005). High dose supplemental vitamin D reduced fall risk by 19% (pooled relative risk (RR) 0.81, 95% CI 0.71 to 0.92; n=1921 from seven trials), whereas achieved serum 25(OH)D concentrations of 60 nmol/l or more resulted in a 23% fall reduction (pooled RR 0.77, 95% CI 0.65 to 0.90). Falls were not notably reduced by low dose supplemental vitamin D (pooled RR 1.10, 95% CI 0.89 to 1.35; n=505 from two trials) or by achieved serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l (pooled RR 1.35, 95% CI 0.98 to 1.84). Two randomised controlled trials (n=624) of active forms of vitamin D met our inclusion criteria. Active forms of vitamin D reduced fall risk by 22% (pooled RR 0.78, 95% CI 0.64 to 0.94).

**Conclusions** Supplemental vitamin D in a dose of 700-1000 IU a day reduced the risk of falling among older individuals by 19% and to a similar degree as active forms of vitamin D. Doses of supplemental vitamin D of less than 700 IU or serum 25-hydroxyvitamin D concentrations of

less than 60 nmol/l may not reduce the risk of falling among older individuals.

## INTRODUCTION

Each year one in three people aged 65 years or older experiences at least one fall,<sup>1-3</sup> with 9% of falls leading to an emergency room visit and 5-6% resulting in a fracture.<sup>4</sup> Fall prevention has, therefore, become a public health goal, especially as the older proportion of the population grows.

Vitamin D has direct effects on muscle strength modulated by specific vitamin D receptors present in human muscle tissue.<sup>5,6</sup> Myopathy from severe vitamin D deficiency presents as muscle weakness and pain,<sup>7</sup> but is reversible with vitamin D supplementation.<sup>8</sup> In several trials of older individuals at risk for vitamin D deficiency, vitamin D supplementation improved strength, function, and balance in a dose-related pattern.<sup>4,8,9</sup> Most importantly, these benefits translated into a reduction in falls.<sup>4,8,9</sup>

Overall, however, results have been mixed for fall prevention with vitamin D; for example, several trials of vitamin D have had non-significant results. This may be explained in part by the use of low doses of vitamin D, as suggested by a 2004 meta-analysis of limited data from three trials on supplemental vitamin D.<sup>10</sup> Other potential explanations include the availability of vitamin D over the counter for the control group; the use of an open trial design, which biases trial results towards the null;<sup>11</sup> and incomplete assessment, inadequate definition, or incomplete ascertainment of falls during the entire observation period,<sup>12</sup> again introducing a bias towards the null.

Several trials on vitamin D have been performed since 2004; thus, the importance of vitamin D dose for the prevention of falls should be reassessed. Specifically, we need to establish the optimum threshold of serum 25-hydroxyvitamin D<sub>3</sub> (25(OH)D; calcidiol (25-hydroxycholecalciferol)) required to prevent falls in older individuals. Notably, two epidemiology studies among older individuals have found a dose-response relationship between lower extremity function and serum 25(OH)D concentrations,<sup>13,14</sup> with

# Last but not least...

- ▶ “A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system.”
  - Evidence based medicine and the medical curriculum (BMJ 2008;337:704–705)

# Lecture 6 of Y1 C3

## Statistical Packages Interpretation of Computer Outputs

# Statistical Packages

- ▶ Many Open Source and Proprietary software are available
  - [https://en.wikipedia.org/wiki/List\\_of\\_statistical\\_packages](https://en.wikipedia.org/wiki/List_of_statistical_packages)
- ▶ Most popular ones:
  - R, SAS, SPSS, Stata are comprehensive statistics packages
  - MATLAB, Maple, Mathematica have statistical features
  - Microsoft excel also has some statistical features in its functions and Data Analysis tools

# SPSS

- ▶ Descriptive statistics
  - For qualitative variables
    - Frequency analysis
  - For quantitative variables
    - Descriptive measures (mean, std dev, etc.)
- ▶ Descriptive statistics for grouped data
- ▶ Transform a variable
- ▶ Normality test
- ▶ Comparison tests
  - T-tests; Sign, Mann Whitney u, Wilcoxon
  - ANOVA; Kruskal-Wallis
- ▶ Crosstabs and Chi-Square tests
- ▶ Correlation analysis
- ▶ Regression analysis
- ▶ Odds Ratio and Relative Risk