ATLAS
ASIAN CLINICAL TRIALS NETWORK FOR CANCERS PROJECT

National Cancer Center Japan

Describing Data

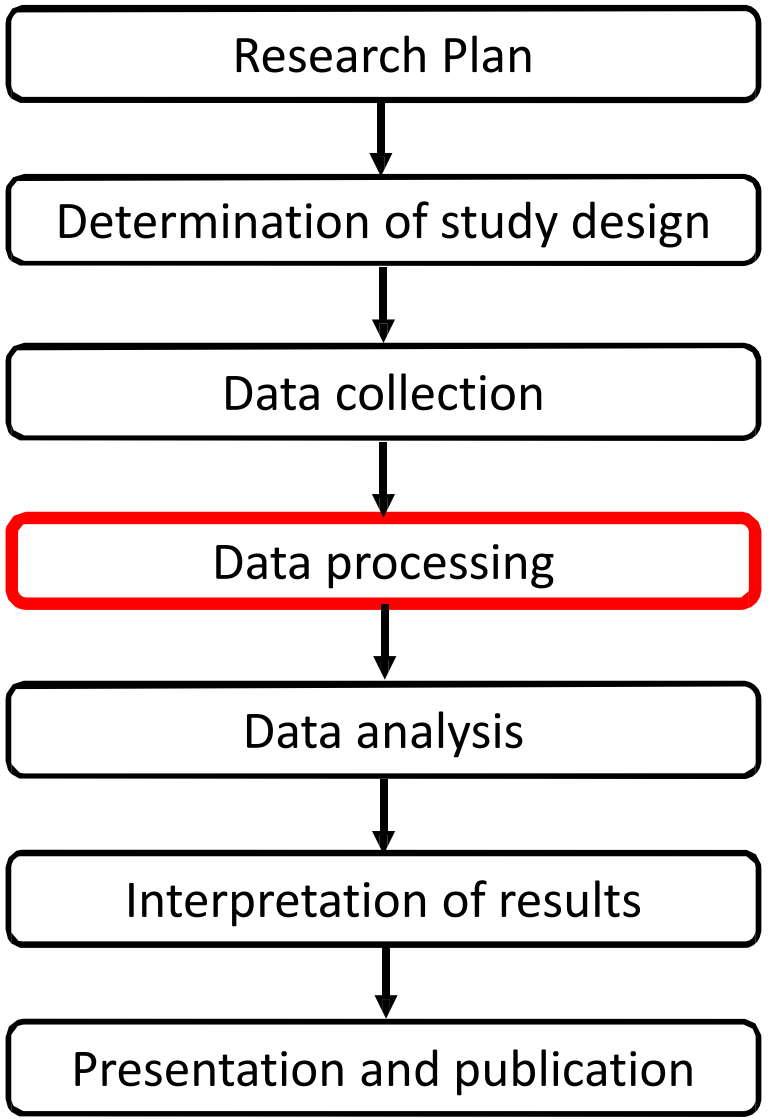
National Cancer Center, Japan

Biostatistics Division/JCOG Data Center

Junki Mizusawa

JCOG
Japan Clinical Oncology Group

General Research Flow



What is your clinical hypothesis?

Prospective? Retrospective? Single group? Randomized?

Which data on which patients are collected?

Any unusual data? How do we present the data?

Are there differences between groups? Is XX a prognostic factor?

Clinical hypothesis addressed? What are the limitations of the study?

What is the next study?

Lecture Content

- Precautions before analyzing data
  - Be suspicious of the entered data
  - Traps in Excel data
- Summarizing continuous variable data with one variable
  - Histogram
  - Summary statistics
- Summarizing continuous variable data with two variables
  - Scatter plots and correlation coefficients
  - Cutoff for continuous variable data
- How to summarize categorical data
  - Contingency table
- Analysis of population and patient flow diagram

Precautions to Take Before Analyzing Data

Avoiding Common Pitfalls Before Statistical Analysis

Is the Entered Data Correct in the First Place?

- When a researcher enters data into a computer, input errors cannot be 0%.
  - The reported input error per field is 0.5% – 6.9%.
    - Neaton, James D., et al. Statistics in medicine 9.1-2 (1990): 115-124.
    - Norton, Susan L., et al. Computers and Biomedical Research 14.2 (1981): 179-198.
    - Kronmal, Richard A., et al. Computers and Biomedical Research 11.6 (1978): 553-566.
  - A significant difference exists between veteran data managers and ordinary clinicians.
    - by Dr. Fukuda@JCOG DC
- What we do at the data center for clinical study
  - Double entry (two persons enter data separately and check if they match)
  - Reading review (one people enter data and check if the data matches the original data [medical record and CRF])
  - Double entry reduces more entry errors than reading review but takes more time.
    - Kawado, Miyuki, et al. Controlled clinical trials 24.5 (2003): 560-569.

Differences in Operating Systems: Receiving Data in Excel on a Mac

	A	B	C
1	Patient ID	Date of surgery	Date of death
2	1	2010/11/20	2012/5/6
3	2	2010/11/23	2014/3/30
4	3	2010/12/6	2013/2/8
5	4	2011/1/6	2012/7/6
6	5	2011/3/16	2011/6/21
7	6	2011/5/24	2015/3/11
8	7	2011/8/1	2015/1/4
9	8	2011/10/9	2014/4/6
10	9	2011/12/17	2013/9/11
11	10	2012/2/24	2014/10/2

- Data include many dates in the future!
- Data provider confirms that the data was entered correctly.
- Why did this happen?

Macs and Windows Have Different Date Systems

Windows settings

Base date: 1900/1/1 counted as 1

A	B	C
Patient ID	Date of surgery	Date of death
1	2010/11/20	2012/5/6
2	2010/11/23	2014/3/30
3	2010/12/6	2013/2/8
4	2011/1/6	2012/7/6
5	2011/3/16	2011/6/21
6	2011/5/24	2015/3/11
7	2011/8/1	2015/1/4
8	2011/10/9	2014/4/6
9	2011/12/17	2013/9/11
10	2012/2/24	2014/10/2

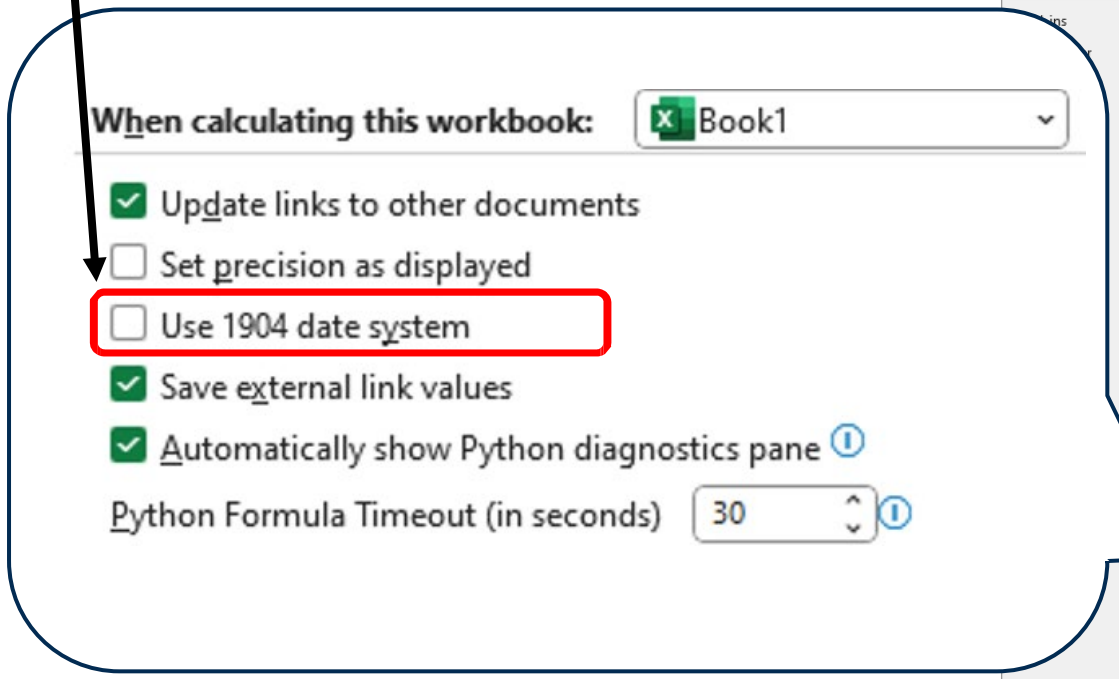
Mac settings

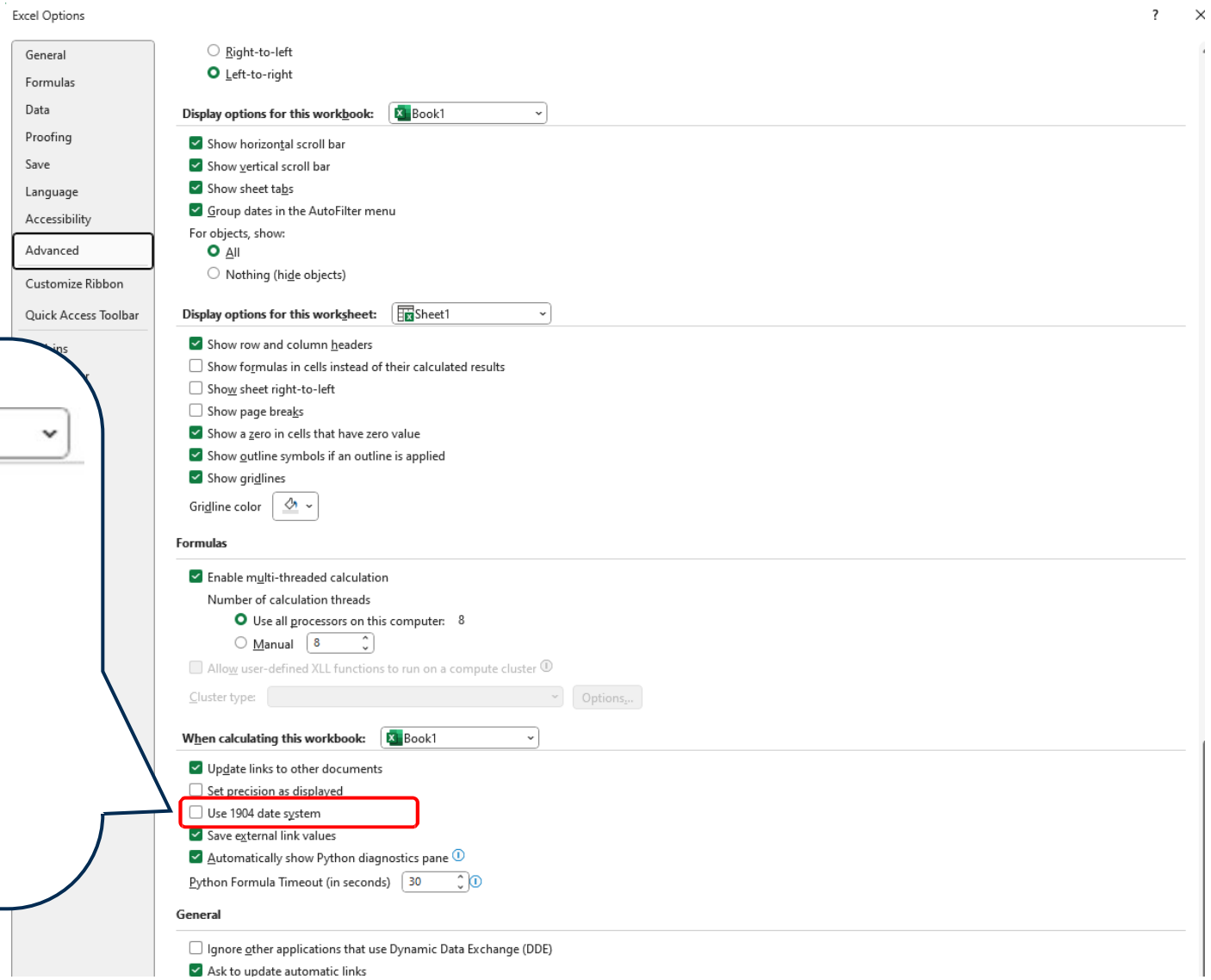
Base date: 1904/1/1 counted as 0

A	B	C
Patient ID	Date of surgery	Date of death
1	2014/11/21	2016/5/7
2	2014/11/24	2018/3/31
3	2014/12/7	2017/2/9
4	2015/1/7	2016/7/7
5	2015/3/17	2015/6/22
6	2015/5/25	2019/3/12
7	2015/8/2	2019/1/5
8	2015/10/10	2018/4/7
9	2015/12/18	2017/9/12
10	2016/2/25	2018/10/3

[File] → [Options] → [Advanced settings]

If this check box is checked, the data set will be Mac-compatible





Summarizing Continuous Variable Data with One Variable

Before performing *t*-test and Wilcoxon rank sum test

Outcome Types and Statistical Methods

	Continuous variable	Binary (0/1)	Survival time
Outcome examples	Blood pressure level Clinical laboratory values	Response rate Adverse events rate	Overall survival Progression-free survival (PFS)
Data summarization methods	Histogram Scatter plot	Contingency table	Kaplan–Meier curve
Group comparison (test)	*t*-test Wilcoxon rank sum test	Chi-square test Fisher's exact test	Log-rank test
Model fitting	Multiple regression analysis	Logistic regression	Cox regression

Visualize and Summarize Data before Testing!

Visualization

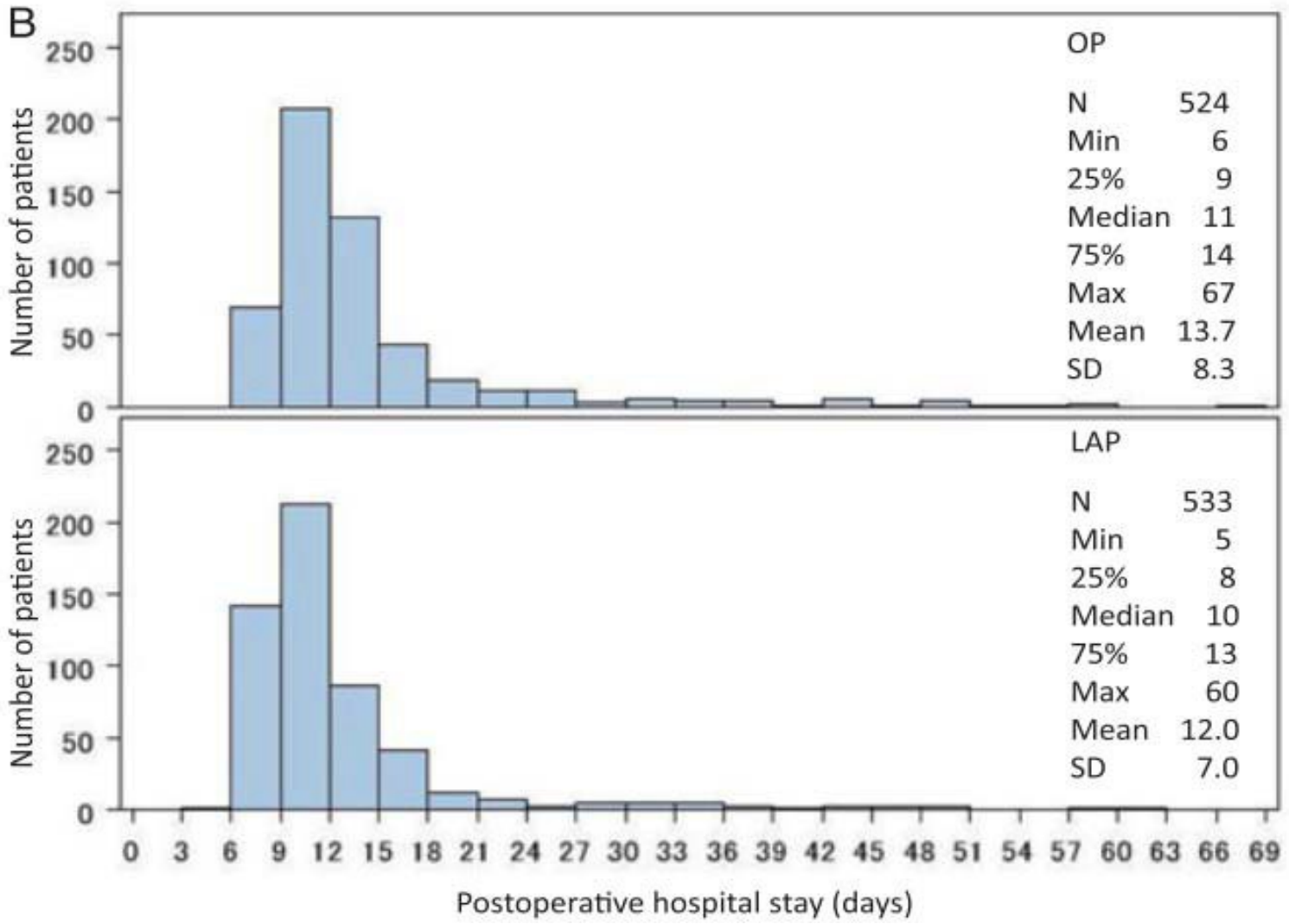
- Graphs
  - Histogram
  - Box plot

Summary statistics

- Location and size of data
  - Mean
  - Median
- Variability and spread of data
  - Standard deviation (SD)
  - First quartile, third quartile
    (Inter quartile range [IQR])
  - Minimum and maximum values (Range)

Histogram: Graphical Representation of One Variable

Length of hospital stay following surgical resection for colorectal cancer



Yamamoto S, et al., *Ann Surg*. 2014; 260(1): 23-30.

Open resection (OP)

Laparoscopic resection (LAP)

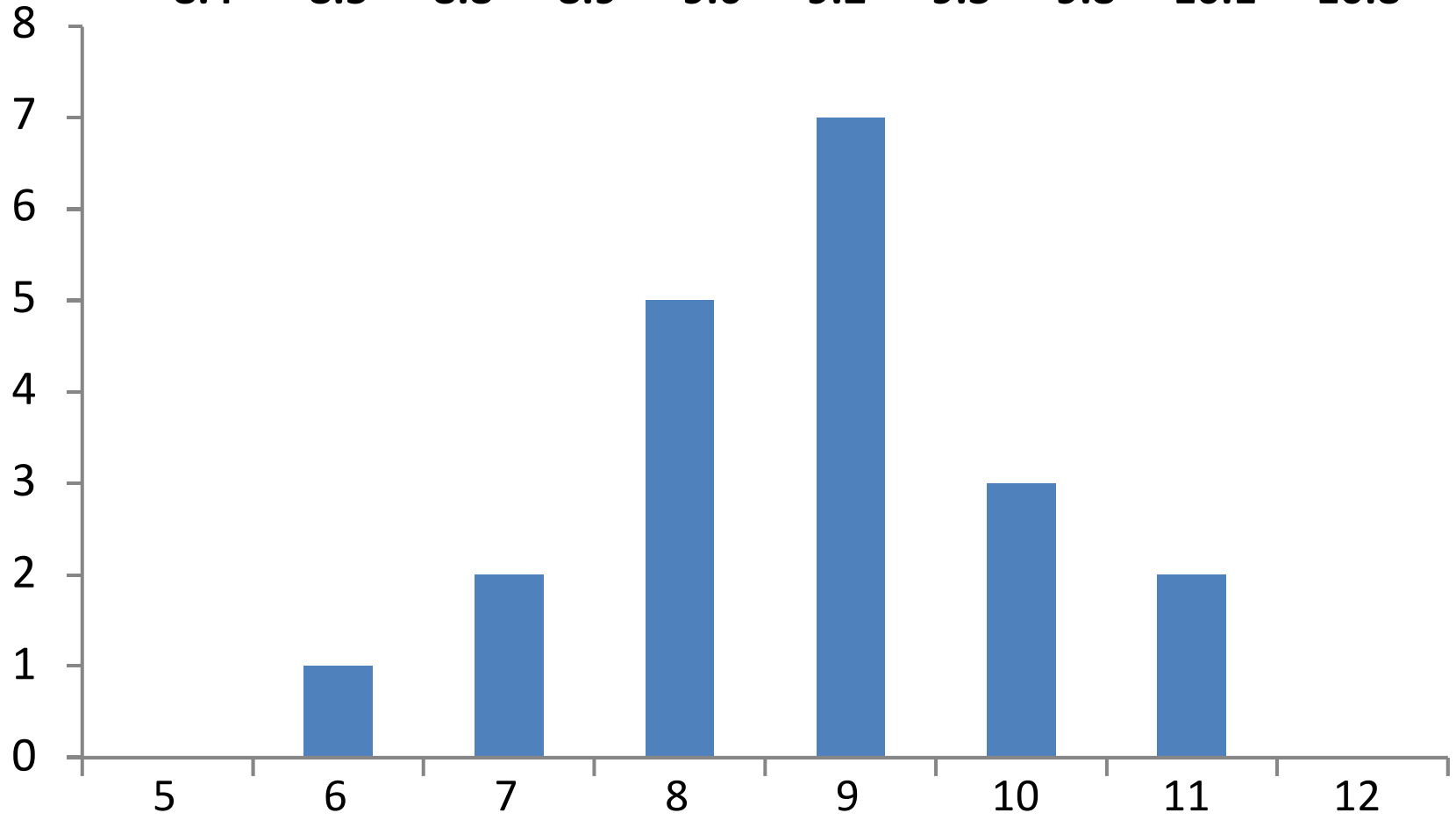
Purpose of Plotting Histograms

- Check for outliers and unusual data points
  - Creatinine value = 0.02, etc.
  - Are there any data that fall outside the eligibility criteria for clinical studies?
- Find out if there is a need for stratification
  - Whether the group exists
  - Whether it is bimodal
- Examine the shape of the distribution
  - Is it symmetrical or does it have a tail in one direction?

Example of Data Points for 20 People

5.9 6.8 7.0 7.2 7.4 7.5 7.8 7.8 8.1 8.3

8.4 8.5 8.8 8.9 9.0 9.2 9.3 9.8 10.1 10.8



Summary Index When Data Follow a Normal Distribution

- Mean

$$\text{Mean Value} = \frac{5.9 + 6.8 + \ldots + 10.1 + 10.8}{20} = 8.33$$

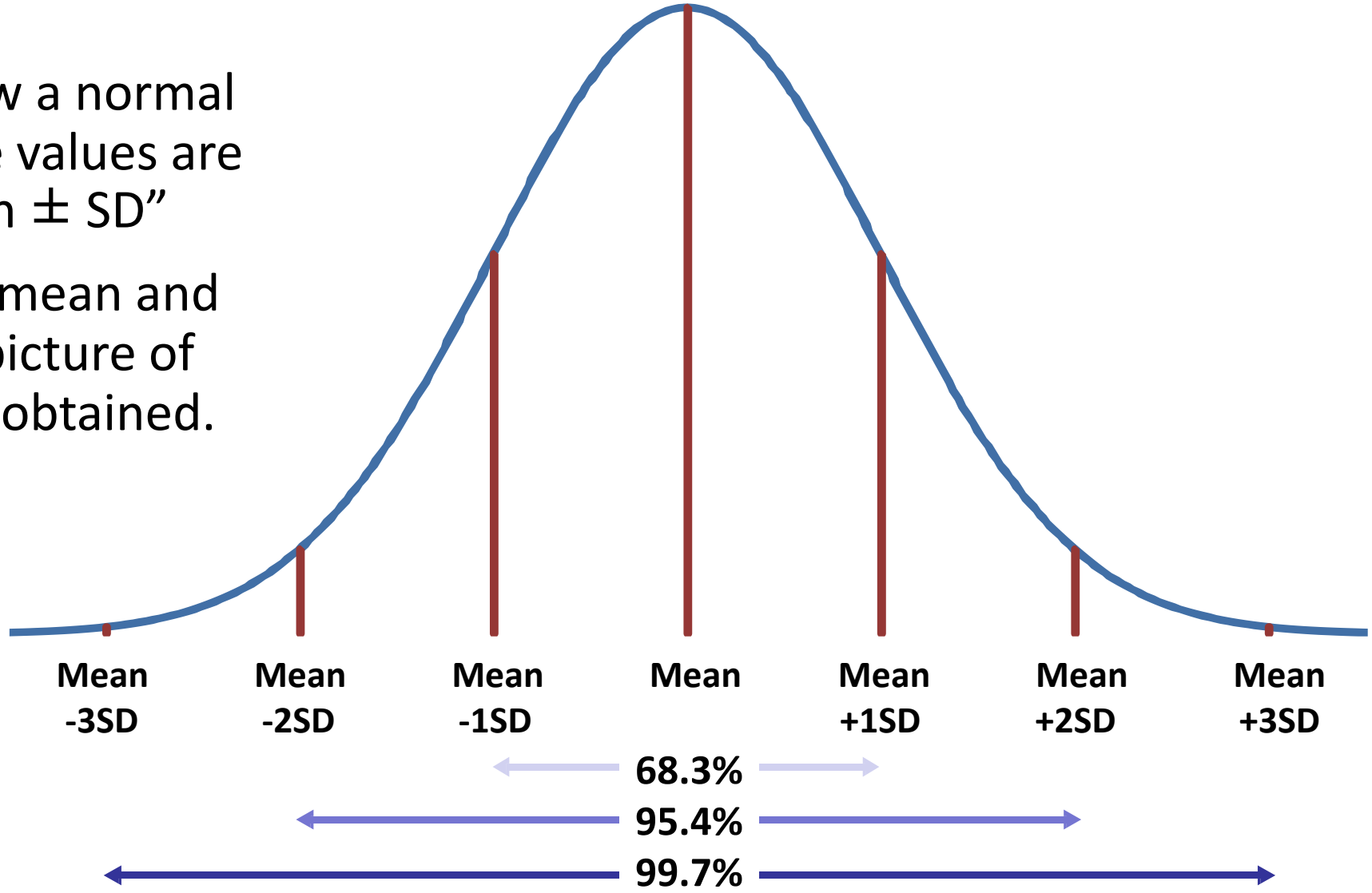
- Standard deviation [SD]
  - Numerator: Variance. The difference between each piece of data and the mean squared and summed.
  - Denominator: Divisor. Number of data – 1
  - The root of the above fraction

$$SD = \sqrt{\frac{(5.9 - 8.33)^2 + (6.8 - 8.33)^2 + \ldots + (10.1 - 8.33)^2 + (10.8 - 8.33)^2}{20 - 1}}$$

$$= 1.13$$

1SD, 2SD, and 3SD Empirical Rules for Normal Distribution

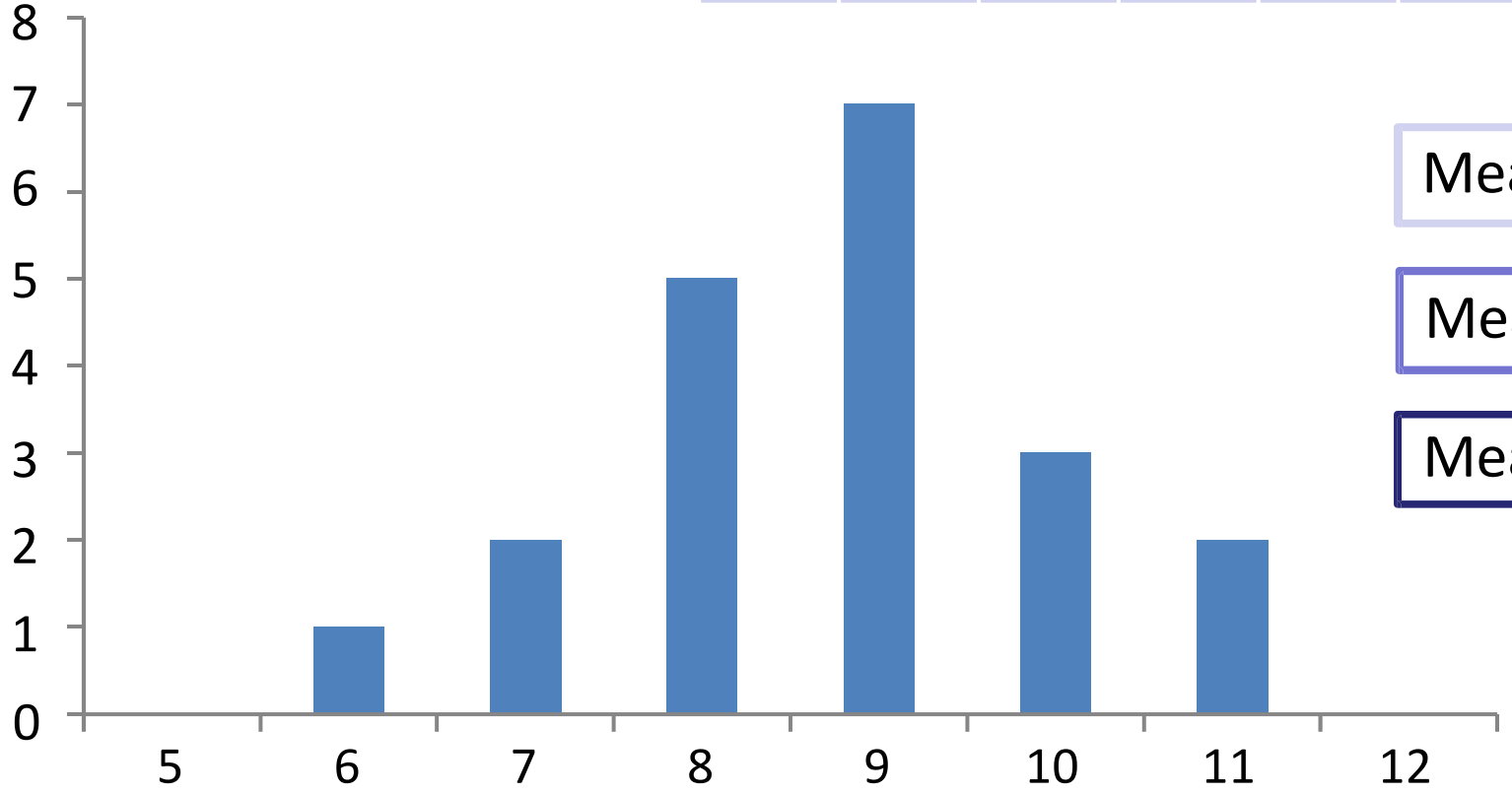
- If the data follow a normal distribution, the values are shown as “Mean ± SD”
- By showing the mean and SD, the overall picture of the data can be obtained.



Example of Data Values for 20 Users

Data for 20 people

5.9	6.8	7.0	7.2	7.4	7.5	7.8	7.8	8.1	8.3
8.4	8.5	8.8	8.9	9.0	9.2	9.3	9.8	10.1	10.8



Mean ± 1 SD: 7.20–9.46

Mean ± 2 SD: 6.09–10.57

Mean ± 3 SD: 4.97–11.69

Interpreting the Mean

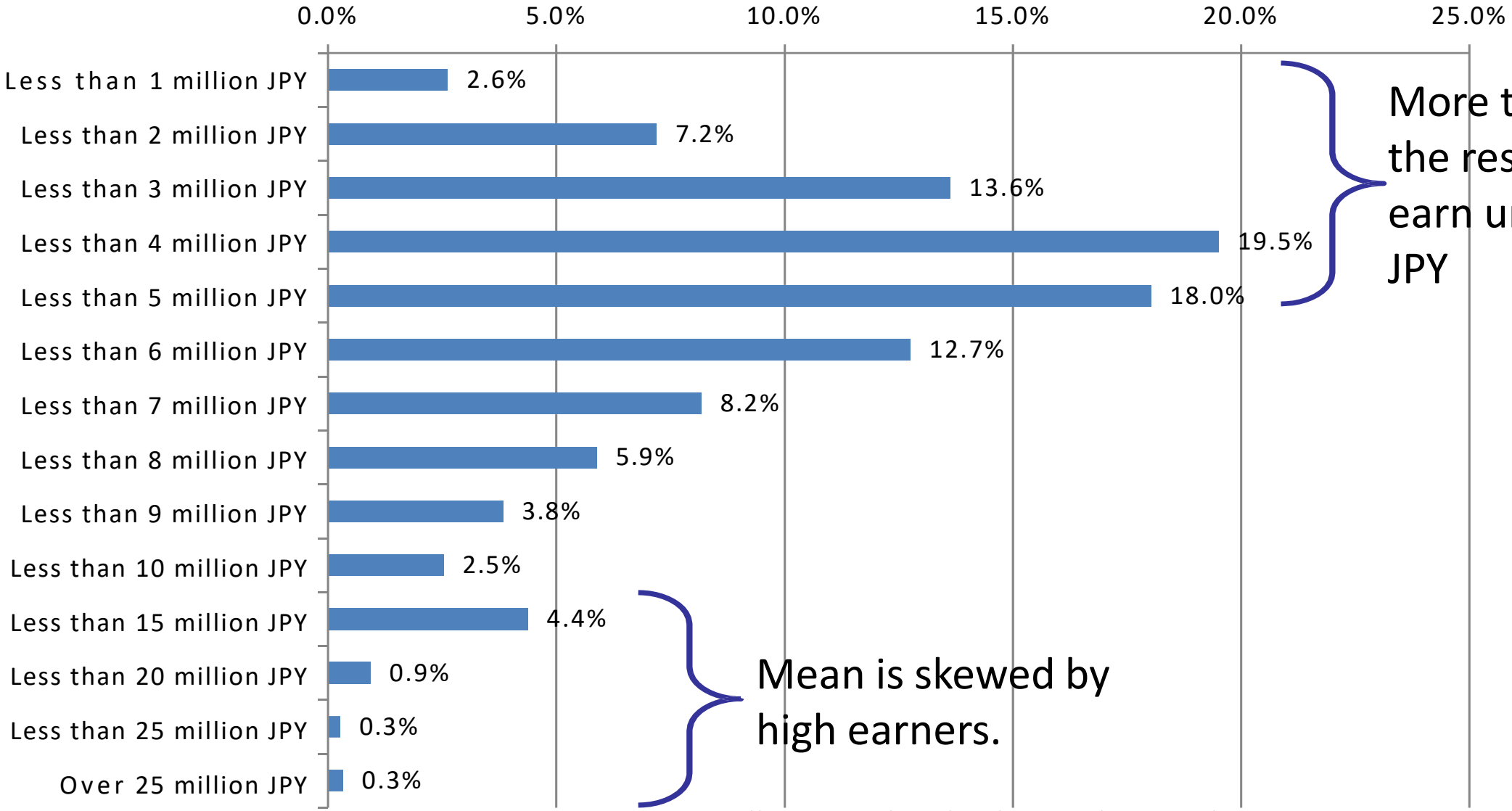
- According to the National Tax Agency's "Survey of actual status for salary in the private sector," the mean annual salary for men was **5.074 million** JPY in FY2010.

- What are the interpretations that can be made from this data?
  - People earning ≤5.07 million JPY and people earning ≥5.07 million JPY each account for 50%
  - More than 50% of the respondents earned <5,070,000 JPY
  - More than 50% of the respondents earned >5,070,000 JPY

Understanding the Characteristics of the Mean

- Numerical example
  - 1, 3, 5, 7, 9 → Mean 5, Median 5
  - 1, 3, 5, 7, **100** → Mean 23.2, Median 5
- The mean is easily influenced by extremely large (small) values compared with other measures
  - Large (small) value = Outlier
- The median always shows the middle value of the data
  - Less susceptible to outliers = robust

The Mean Is Not the Center of the Distribution



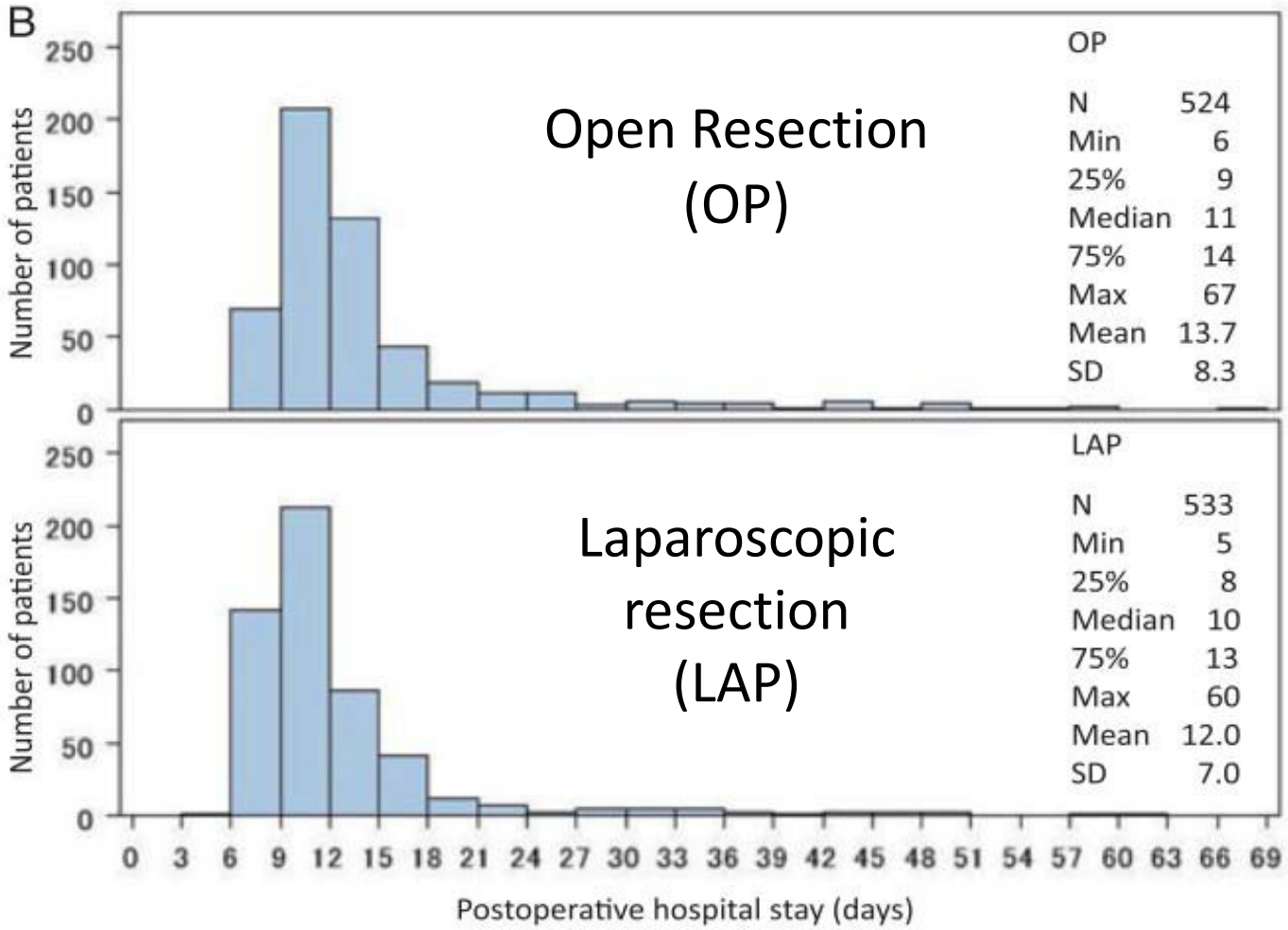
More than 50% of the respondents earn under 5 million JPY

Mean is skewed by high earners.

Created from http://www.nta.go.jp/kohyo/tokei/kokuzeicho/minkan2010/minkan.htm

When the Data Are Not Normally Distributed

Length of hospital stay following surgical resection for colorectal cancer



- The histogram has a tail on the right
- Mean ± SD does not give a complete picture of the data
- Summary indexes other than mean and SD are needed

Yamamoto S, et al., *Annals of Surgery*. 2014; 260(1): 23-30.

Example of Right-Skewed Data Distribution

- Complications after liver resection and background factors

Virani S, et al. *J Am Coll Surg*. 2007; 204(6): 1284-92.

Table 1. Comparison of Comorbidities: Overall and by Morbidity Status

Variable	Overall (n = 783)	With complication (n = 177)	Without complication (n = 606)	p Value
Male gender (%)	48.5	55.4	46.5	0.041
Age (y), mean ± SD	57.5 ± 13.7	58.3 ± 13.8	57.2 ± 13.7	0.340
Caucasian race (%)	80.1	76.8	81.0	0.240
Cardiovascular (%)				
General	**14.2 ± 25.1**	**16.6 ± 24.7**	**13.5 ± 25.1**	
ASA class 3 or higher (%)	54.5	65.0	39.9	0.002
Pack-years of tobacco, mean ± SD	14.2 ± 25.1	16.6 ± 24.7	13.5 ± 25.1	0.167
Smoker (%)	18.4	21.5	17.5	0.227
Alcohol use (%)	4.9	5.7	4.6	0.554
Emergency status (%)	0.8	1.1	0.7	0.622
DNR status (%)	0.0	0.0	0.0	1.000

- In a scenario where only positive values are considered, if the **mean ≤ SD**, the shape of the distribution is considered to be asymmetrical.
- If mean ≤ SD, summarizing the data using the mean and SD is inappropriate.
  - An approximate “Mean > 2SD” serves as a guideline for mean ± SD

Summarizing Data That Does Not Follow a Normal Distribution (1)

- Median
  - Sort by the size of the data
    - If there is an odd number of data, the value of the data in the middle
    - If there is an even number of data, the mean of the two data in the middle

$$\text{Median} = \frac{8.3 + 8.4}{2} = 8.35$$

- First quartile (25% point)
  - The value that separates the lowest 25% of the data from the highest 75% when the data are sorted in ascending order
- Third quartile (75% point)
  - The value that separates the lowest 75% of the data from the highest 25% when the data are sorted in ascending order

$$\text{First quartile} = \frac{7.4 + 7.5}{2} = 7.45; \text{ Third quartile} = \frac{9.0 + 9.2}{2} = 9.10$$

Summarizing Data Than Does Not Follow a Normal Distribution (2)

- Interquartile range (IQR)
  - First quartile (25% point) to third quartile (75% point)
  - Displays the range where half of the data are scattered

- Range
  - Minimum to maximum
  - Displays the full range of data

Presenting Data That Does Not Follow a Normal Distribution

TABLE 3. Operative Results

	Open Surgery (n = 524)	Laparoscopic Surgery (n = 533)	*P*
Procedures			
Ileocecal resection	54 (10.3%)	37 (6.9%)	
Right hemicolectomy	100 (19.1%)	110 (20.6%)	
Sigmoidectomy	208 (39.7%)	240 (45.0%)	
Anterior resection	155 (29.6%)	137 (25.7%)	0.1534
Partial resection	4 (0.8%)	7 (1.3%)	
Hartmann	1 (0.2%)	0	
Others	2 (0.4%)	2 (0.4%)	
Lymphadenectomy			
D3	521 (99.4%)	529 (99.2%)	1
Estimated blood loss, mL			
Median	85	30	<0.001
Range	0–3395	0–4080	
Interquartile range	49.5–180	10–70	
Operative time, min			
Median	159	211	<0.001
Range	68–710	80–616	
Interquartile range	130–188.5	179–256	
Blood transfusion within 3 postoperative days			
Yes	6 (1.1%)	4 (0.8%)	0.5437

- It is best to present all three as a set
  - Median
  - Interquartile range
  - Range
- Interquartile ranges are better than ranges for interpreting data variability

Yamamoto S, et al., *Annals of Surgery*. 2014; 260(1): 23-30.

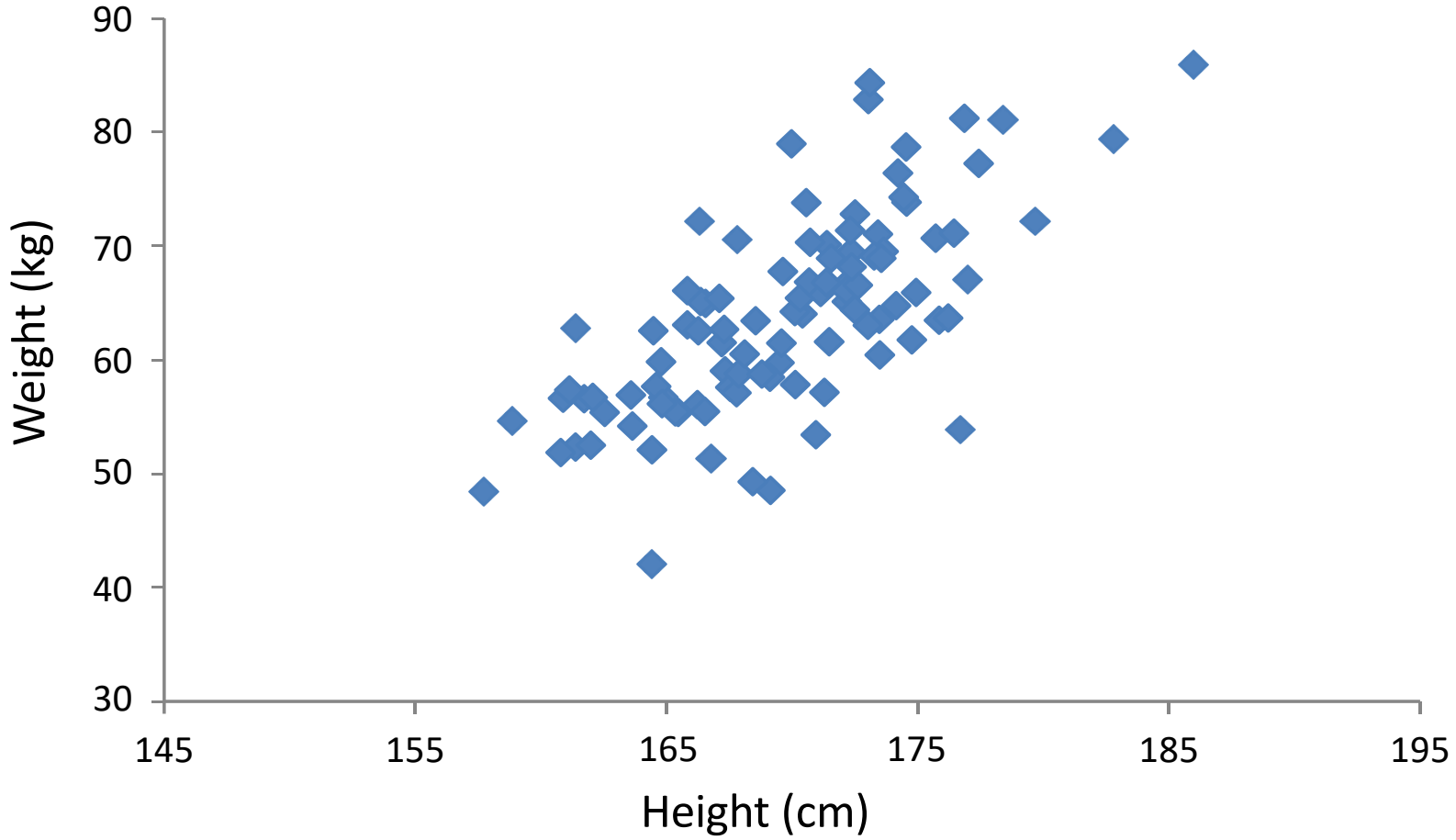
Median Rather Than Mean for Clinical Data

- Clinical data often has a tail to the right
  - Clinical laboratory values
  - Biomarkers
  - Surgery duration, hemorrhage volume
- It is safer to use the median for clinical data
- If you want to test the deviation of a nonsymmetrical distribution, use Wilcoxon rank-sum test instead of *t*-test

Summarizing Data For Two Continuous Variables

Scatter Plot

- Plot of 100 men's height [horizontal axis/cm] and weight [vertical axis/kg]



Purpose of Scatter Plots

- Check for outliners and unusual data
  - 190 cm in height and 45 kg in weight
  - Individual data points look normal but unusual when combined
- Find out if there is a need for stratification
  - Whether the group exists
  - Whether the relationship between the two variables differ for each group
- Understand the approximate degree of correlation
  - Is the correlation linear or curvilinear? How strongly associated?

Correlation

- The original definition was “a **linear** association between two variables of continuous type.”
  - An association is not inherently “correlated” if the relationship is not linear.
    - The phrase “high and low gene expression levels are ‘correlated’ with prognosis” is not an appropriate expression because only two categories have been examined.
- It is better to simply use “associated” unless there is a linear association.

Correlation Coefficient

- A measure of correlation
  - The range is −1 to 1
    - −1: As X increases, Y decreases. Perfect negative correlation
    - 0: No correlation. However, it does not mean “no association.”
    - 1: As X increases, Y increases. Perfect positive correlation.
- Pearson’s product moment correlation coefficient (r)
  - An indicator that measures whether there is a linear relationship between two variables
- Spearman’s rank correlation coefficient (ρ)
  - Measures whether an increase (decrease) in one side will cause an increase (decrease) in the other
    - It is irrelevant whether a linear relationship exists.

Equation for Correlation Coefficient (no need to memorize)

- Pearson's product moment correlation coefficient (r)

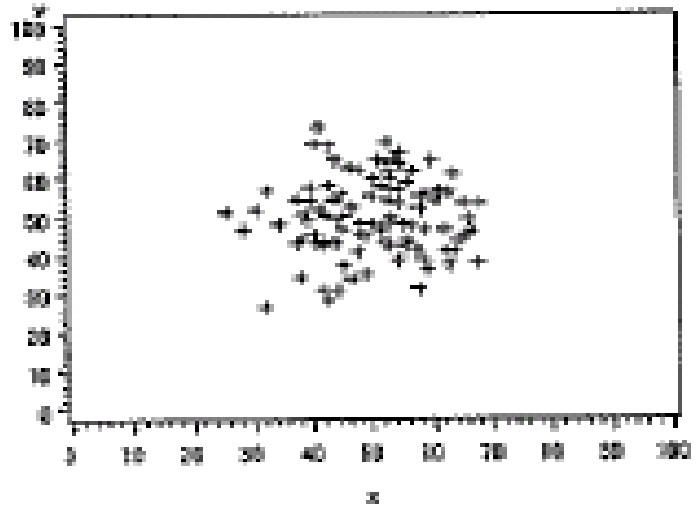
$$r = \frac{\sum_{i=1}^{n} (\text{x value for each data} - \text{mean of x})\ (\text{y-value of each data} - \text{mean of y})}{\sqrt{\left[\sum_{i=1}^{n} (\text{x value of each data} - \text{mean of x})^2\right]\left[\sum_{i=1}^{n} (\text{y-value of each data} - \text{mean of y})^2\right]}}$$

- Spearman's rank correlation coefficient (ρ)
  - Convert all data to rank scale to Pearson's correlation coefficient formula

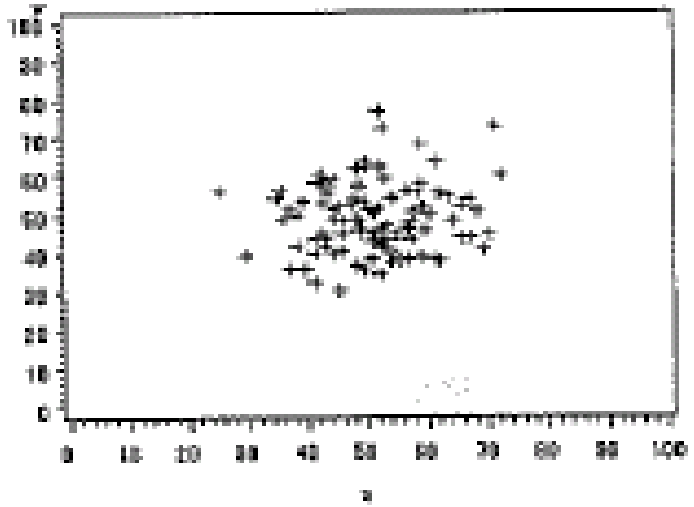
x	y
96	268
86	318
101	289
100	301
98	298

x	y
2	1
1	5
5	2
4	4
3	3

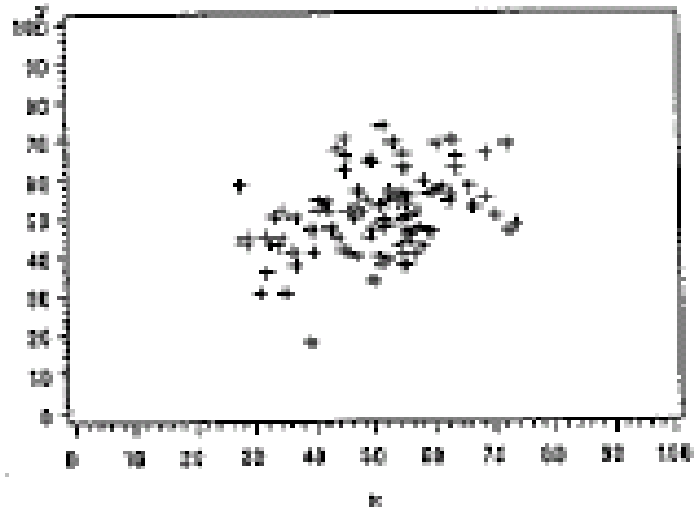
Pearson’s Correlation Coefficient and Scatter Plots (1)



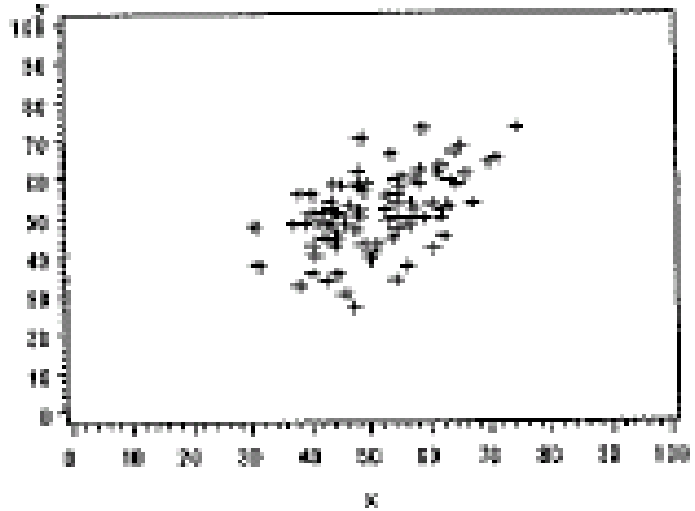
r=0.0



r=0.2



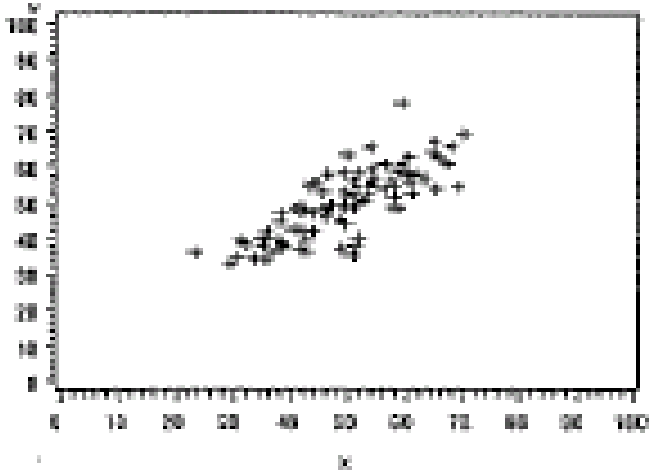
r=0.4



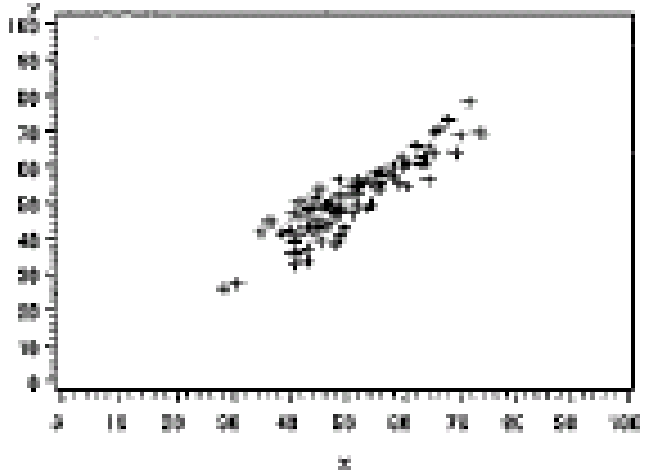
r=0.6

Pearson’s Correlation Coefficient and Scatter Plots (2)

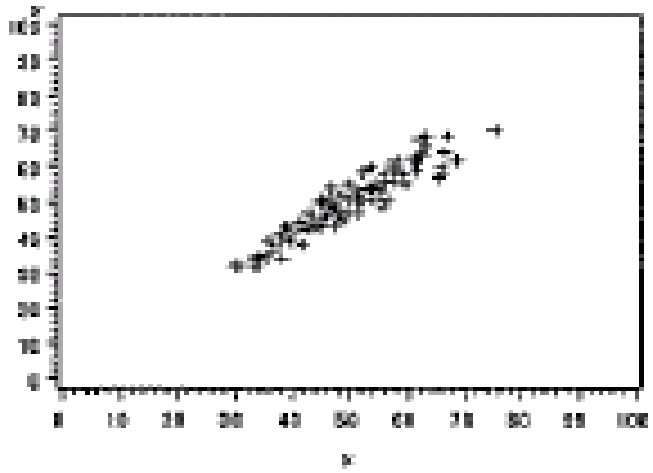
A correlation coefficient of 0.8 does not appear to be correlated.



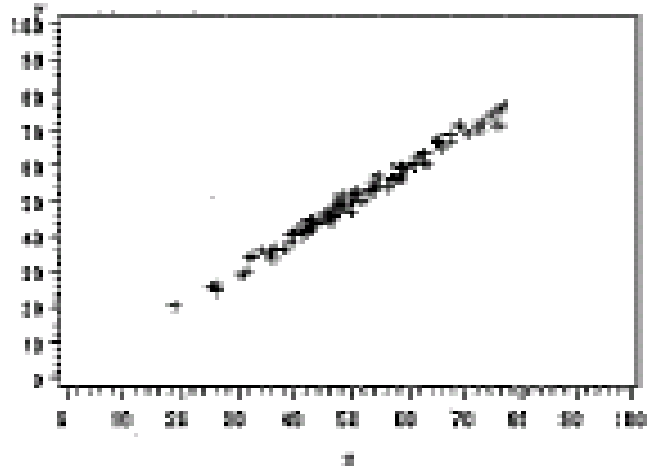
r=0.8



r=0.90



r=0.95

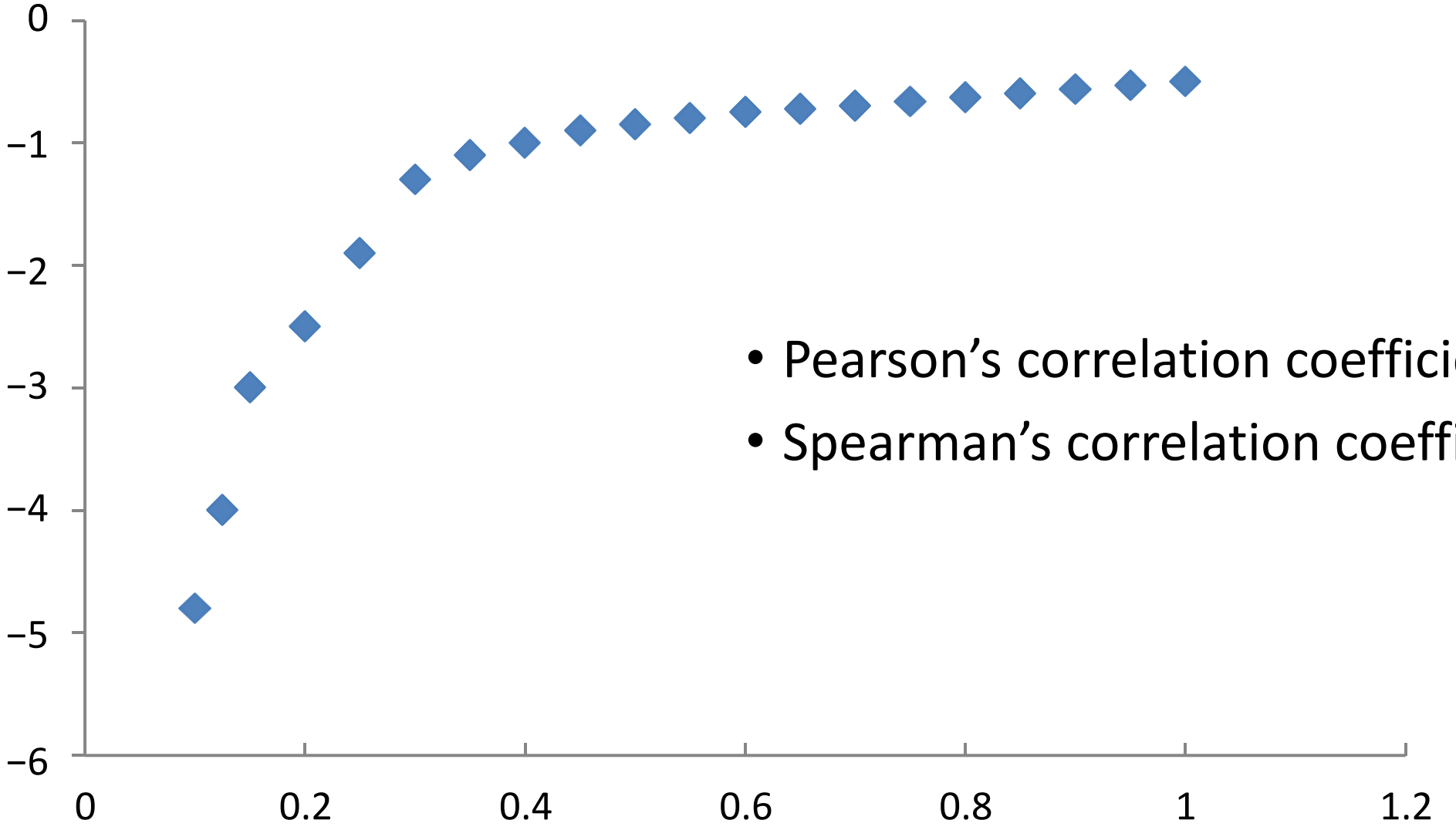


r=0.99

Note on Correlation Coefficients

1. Pearson’s correlation coefficient indicates the strength of the linear association between two variables.
2. Independent of the unit of measurement or appearance.
3. Pearson’s correlation coefficient is sensitive to outliers.
4. Note the presence of groups.
5. Significant differences in correlation coefficients depend on N.

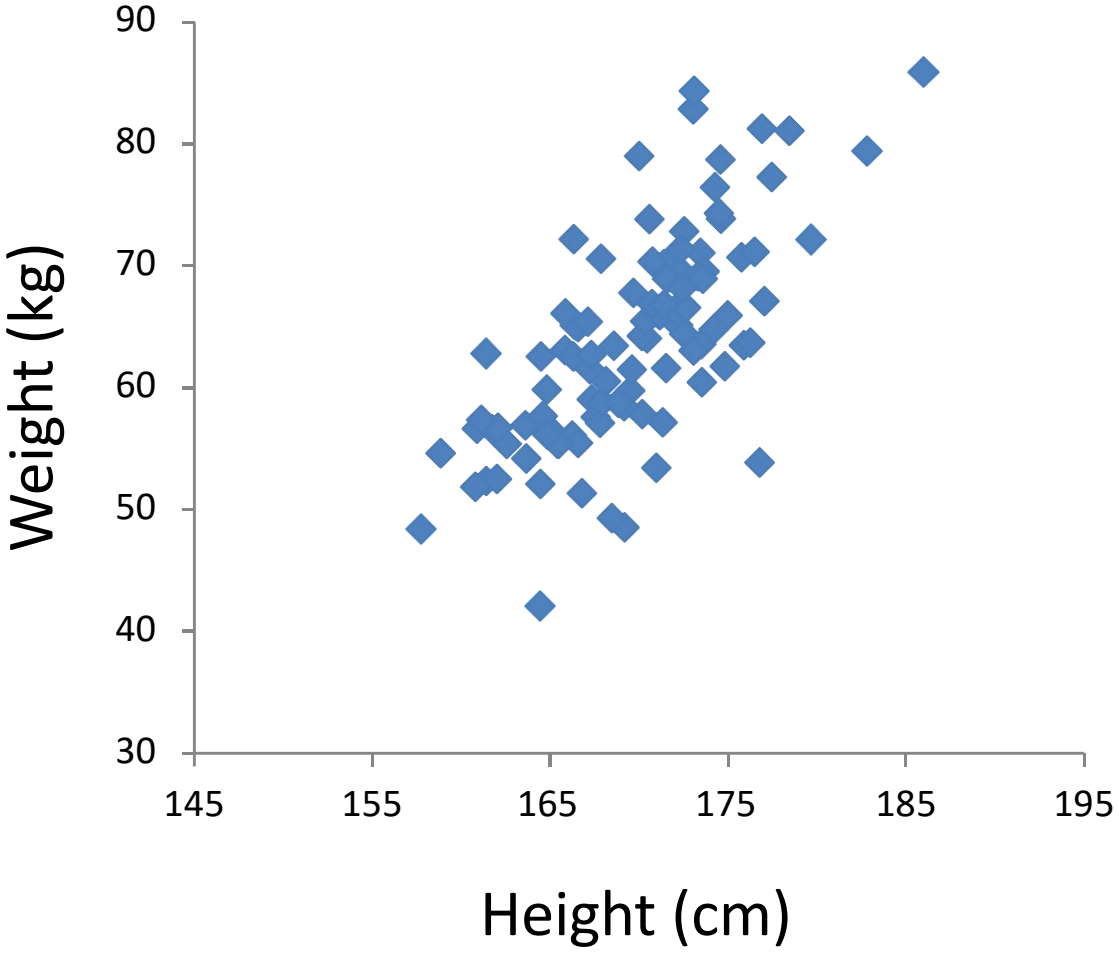
Not a Linear, but There is a Two-Variable Relationship

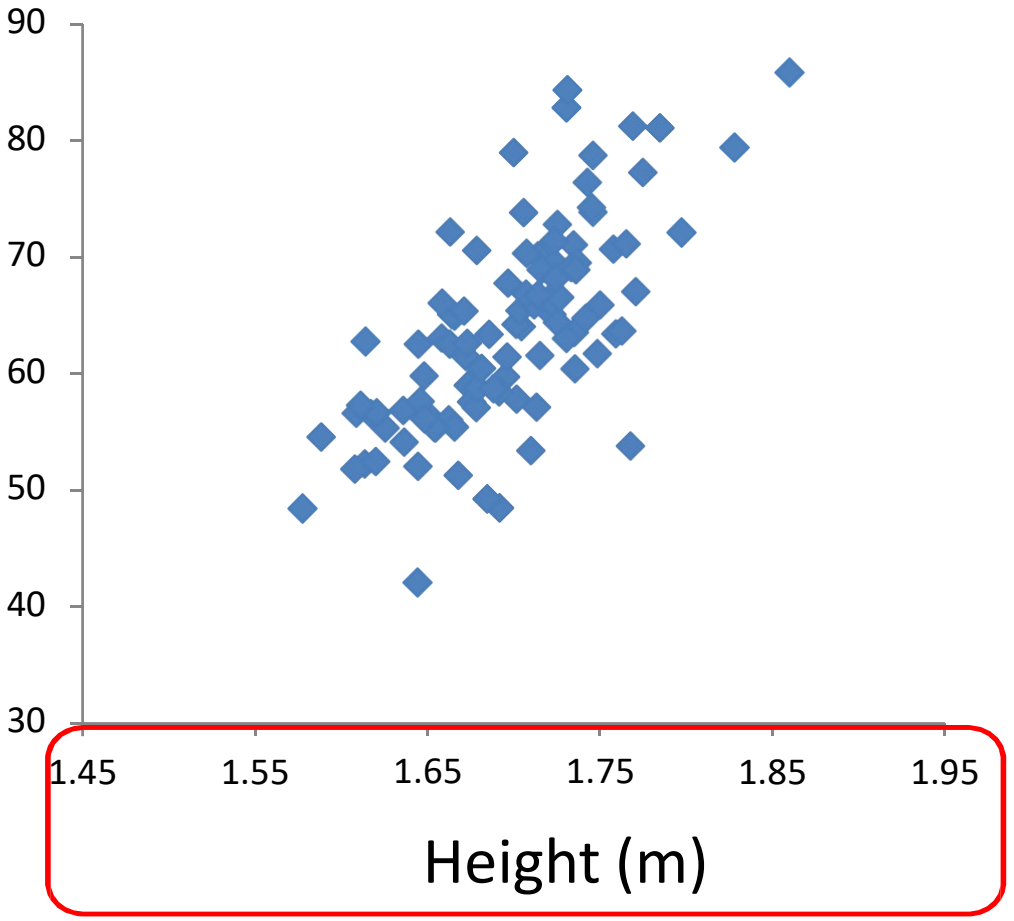


- Pearson's correlation coefficient (r) = 0.77
- Spearman's correlation coefficient (ρ) = 1

Not Dependent on Measurement Unit or Appearance (1)

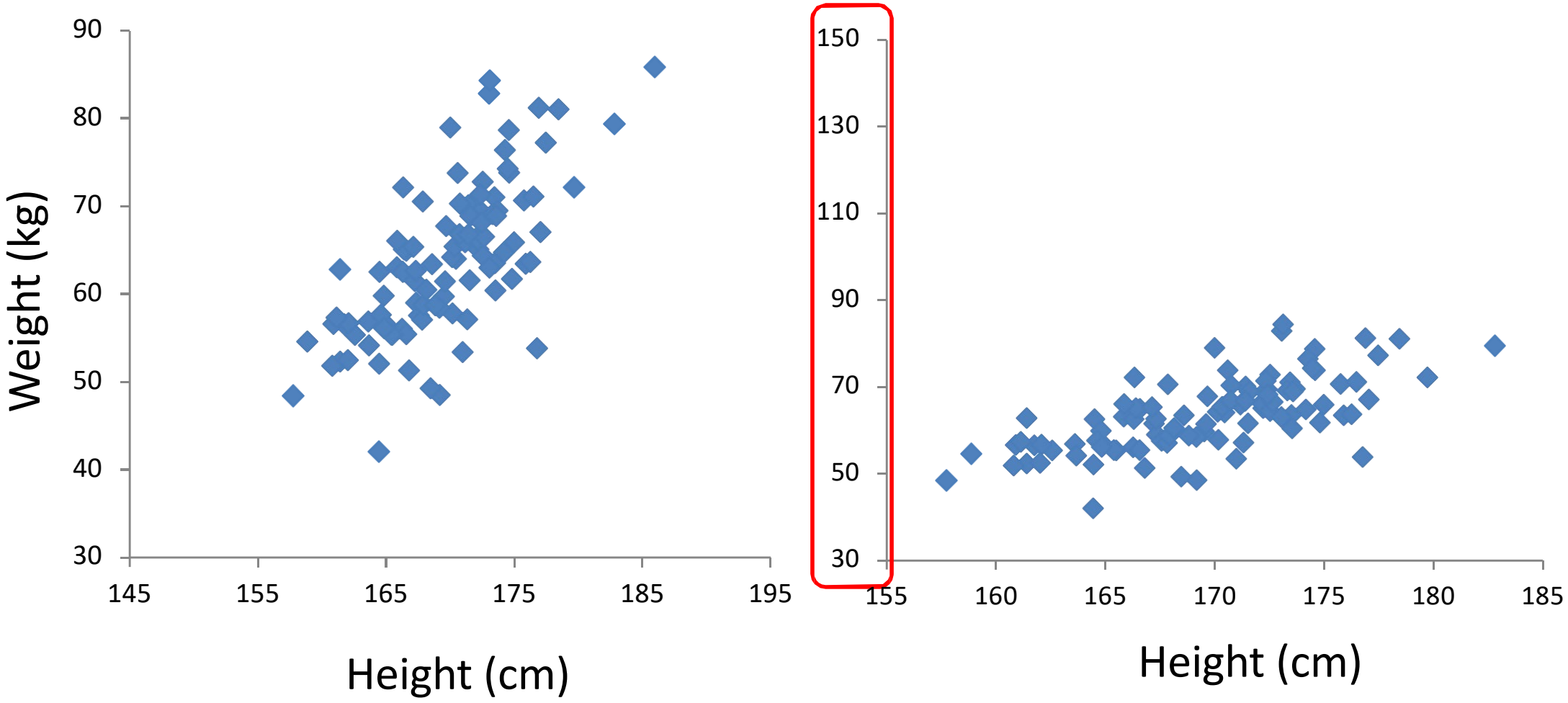
In both cases, Pearson's correlation coefficient (r) = 0.70





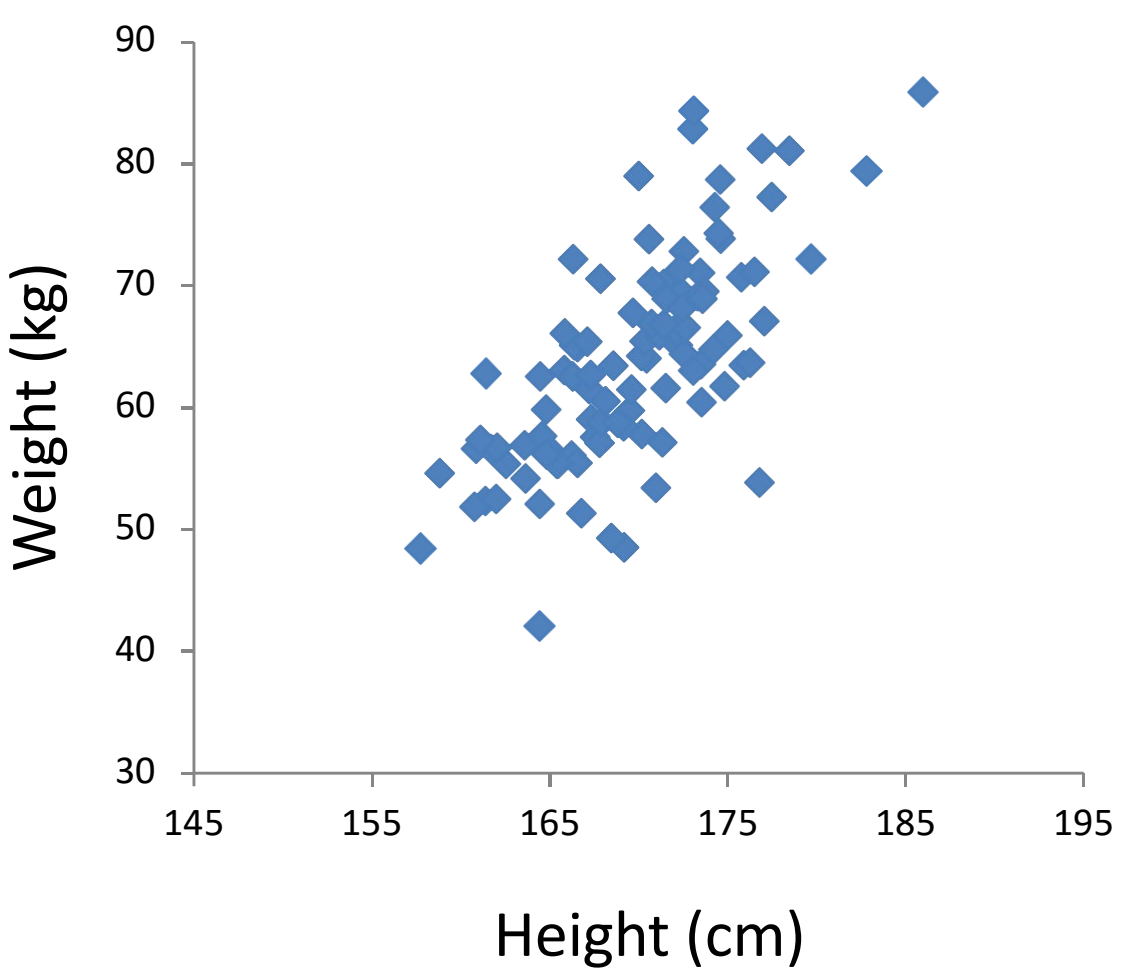
Not Dependent on Measurement Unit or Appearance (2)

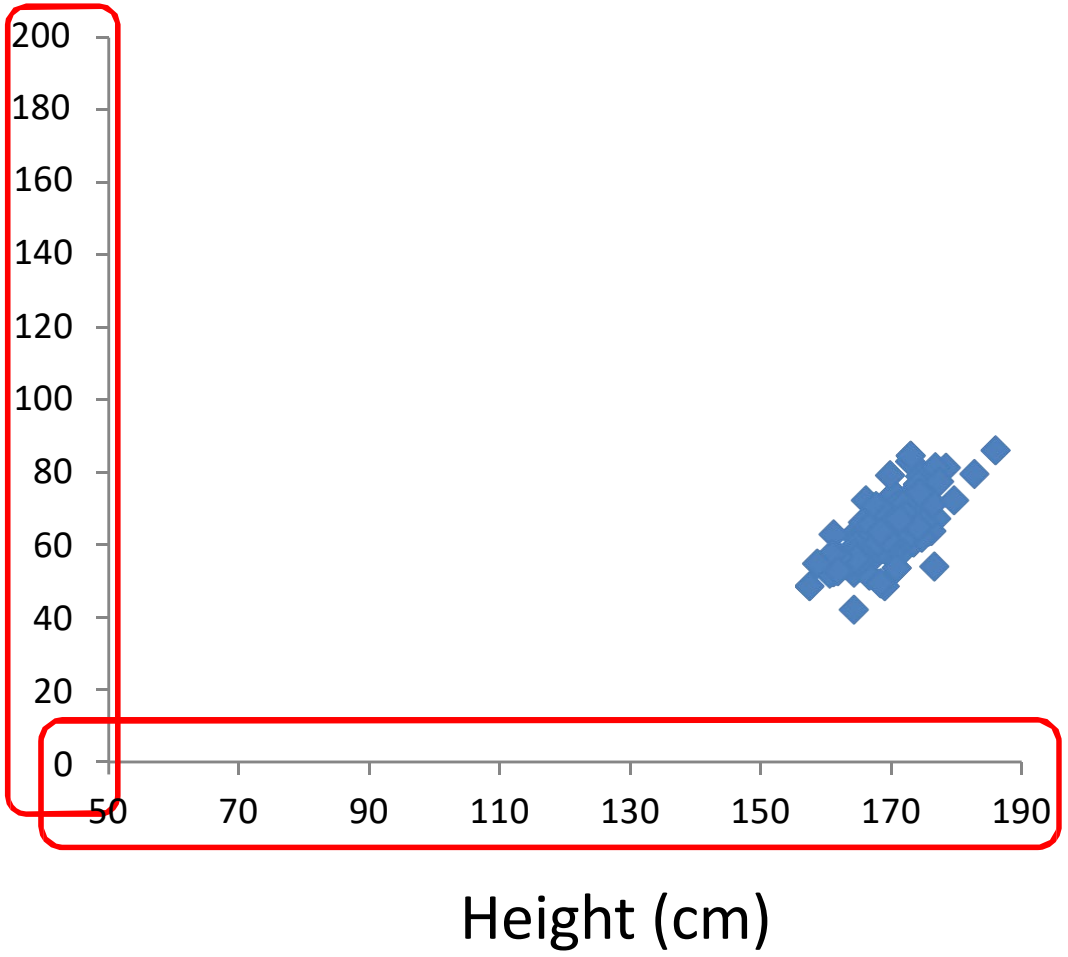
In both cases, Pearson's correlation coefficient (r) = 0.70



Not Dependent on Measurement Unit or Appearance (3)

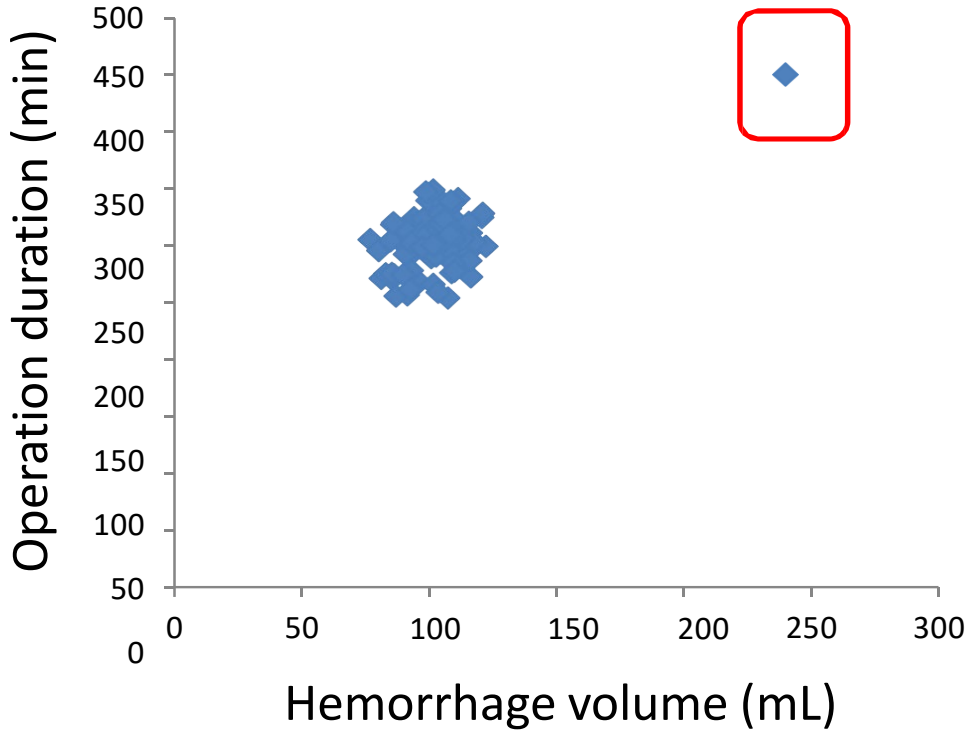
In both cases, Pearson's correlation coefficient (r) = 0.70



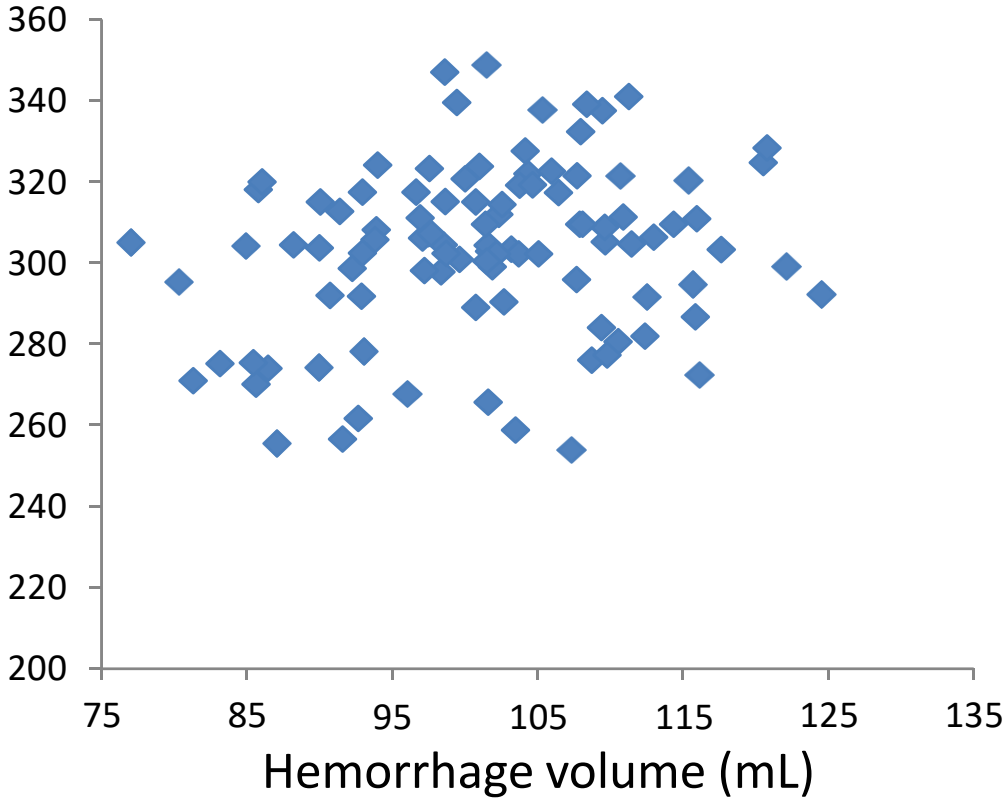


Pearson's Correlation Coefficient Is Sensitice to Outliners

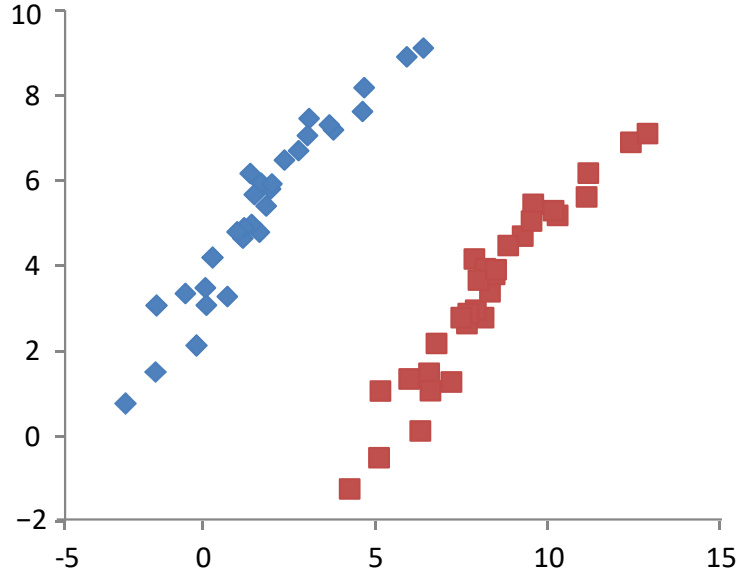
- When including outliners
  - Pearson's correlation coefficient (r) = **0.58**
  - Spearman's correlation coefficient (ρ) = 0.23



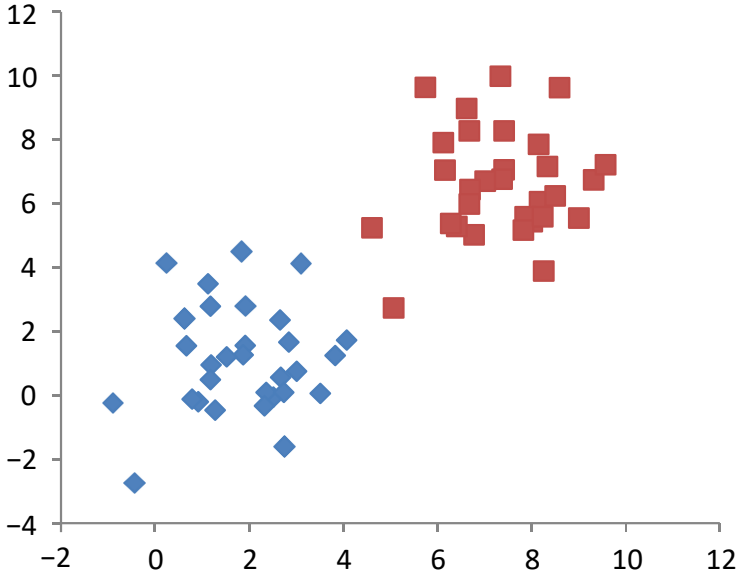
- When including outliners
  - Pearson's correlation coefficient (r) = **0.23**
  - Spearman's correlation coefficient (ρ) = 0.21



Check for Groups in the Data



- Pearson's correlation coefficient (r) by group
  - −0.96 and 0.96
- Pearson's correlation coefficient (r) when combined
  - −0.09



- Pearson's correlation coefficient (r) by group
  - −0.1 and 0.1
- Pearson's correlation coefficient (r) when combined
  - −0.81

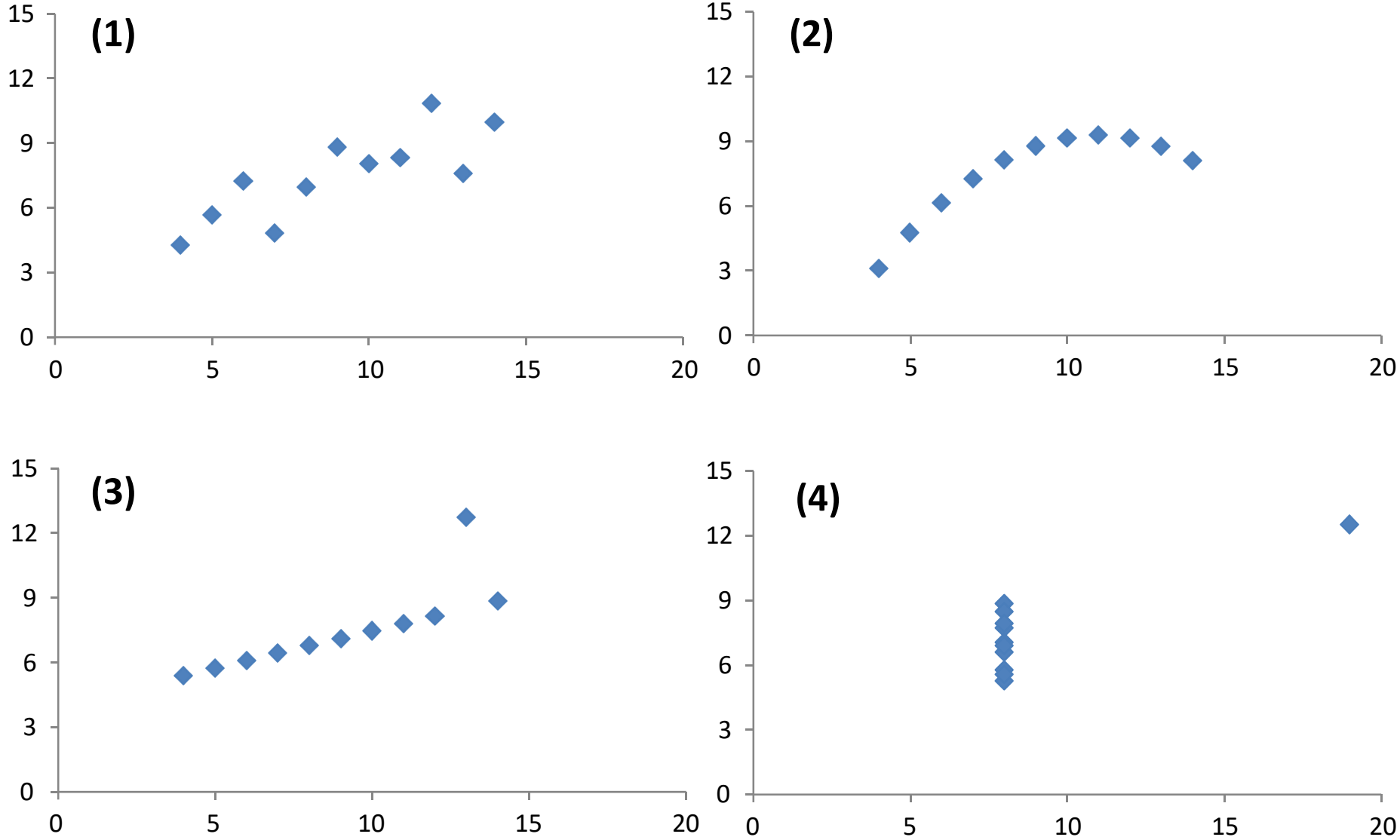
Which Pattern Has the Strongest Correlation?

(1)		(2)		(3)		(4)	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
4	4.26	4	3.1	4	5.39	8	6.58
5	5.68	5	4.74	5	5.73	8	5.76
6	7.24	6	6.13	6	6.08	8	7.71
7	4.82	7	7.26	7	6.42	8	8.84
8	6.95	8	8.14	8	6.77	8	8.47
9	8.81	9	8.77	9	7.11	8	7.04
10	8.04	10	9.14	10	7.46	8	5.25
11	8.33	11	9.26	11	7.81	8	5.56
12	10.84	12	9.13	12	8.15	8	7.91
13	7.58	13	8.74	**13**	**12.74**	8	6.89
14	9.96	14	8.1	14	8.84	**19**	**12.5**

Graphs and Correlation Coefficients

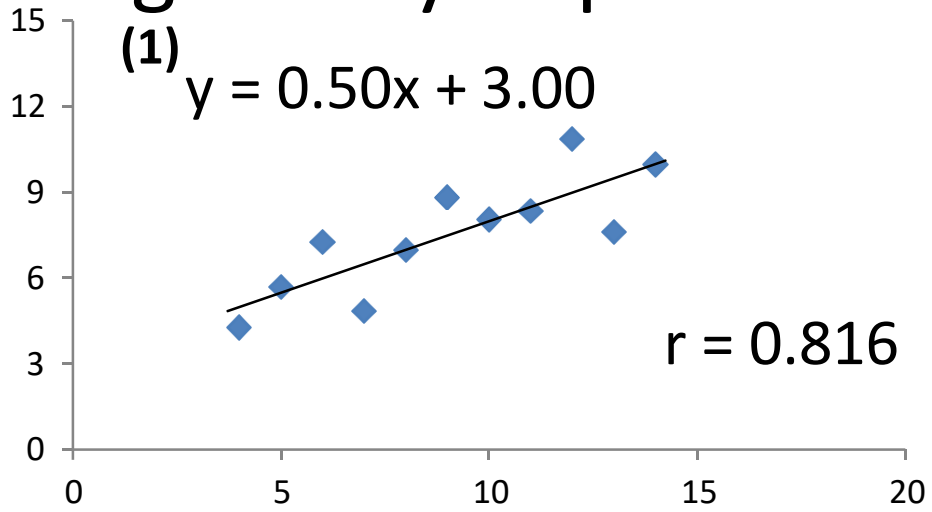
	N	Mean	Standard deviation (SD)	Correlation coefficient (*P* value)	Regression Equation (Relational equation between Y and X)
X_1	11	9.00	3.32	0.816 (p = 0.0022)	$Y_1 = 3.00 + 0.50X_1$
Y_1	11	7.50	2.03		
X_2	11	9.00	3.32	0.816 (p = 0.0022)	$Y_2 = 3.00 + 0.50X_2$
Y_2	11	7.50	2.03		
X_3	11	9.00	3.32	0.816 (p = 0.0022)	$Y_3 = 3.00 + 0.50X_3$
Y_3	11	7.50	2.03		
X_4	11	9.00	3.32	0.816 (p = 0.0022)	$Y_4 = 3.00 + 0.50X_4$
Y_4	11	7.50	2.03		

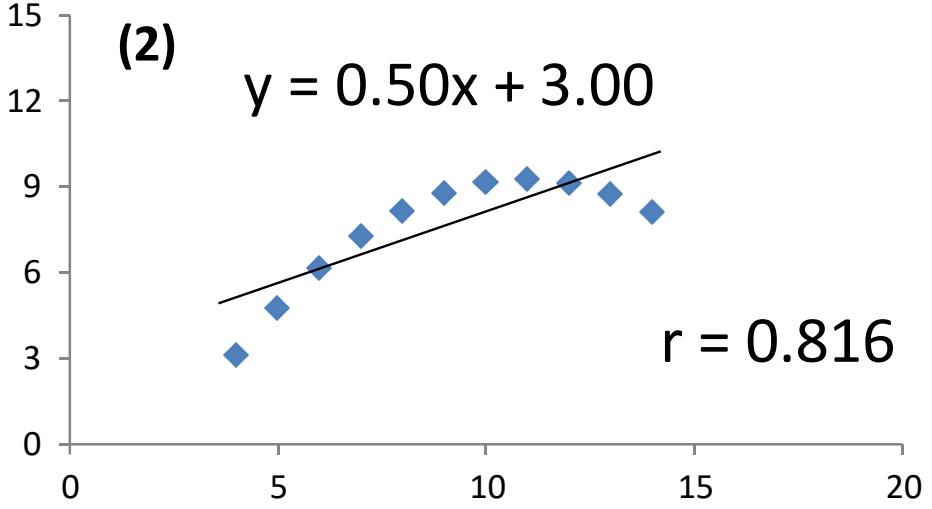
Scatter Plot Results (horizontal axis x, vertical axis y)

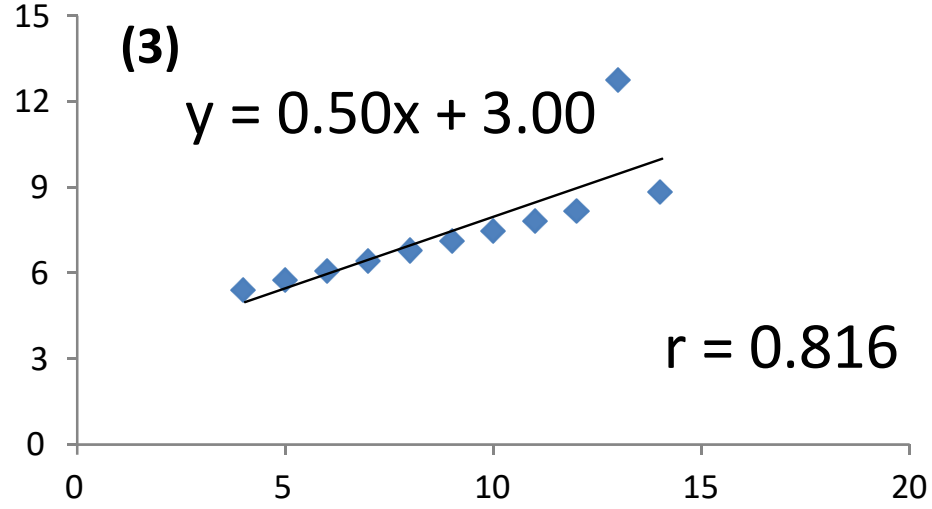


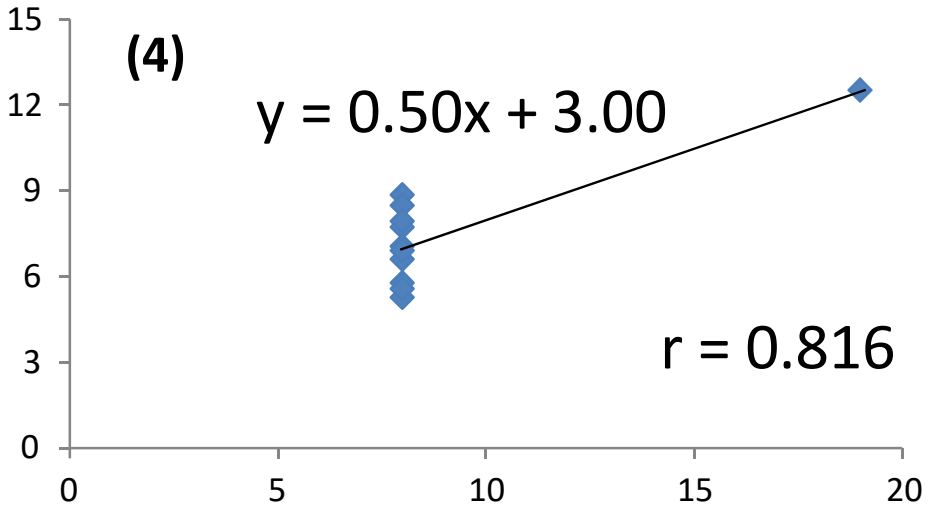
Anscombe, F. J., *American Statistician*, Vol 27, 1973.

Do Not judge Only by Summary Statistics; Graphing is Very Important!









What are the criteria for Determining “Correlation”?

- Because the P value reduce as N increases, it is not always appropriate to judge by statistical significance.
- Decided on a case-by-case basis depending on the research area.

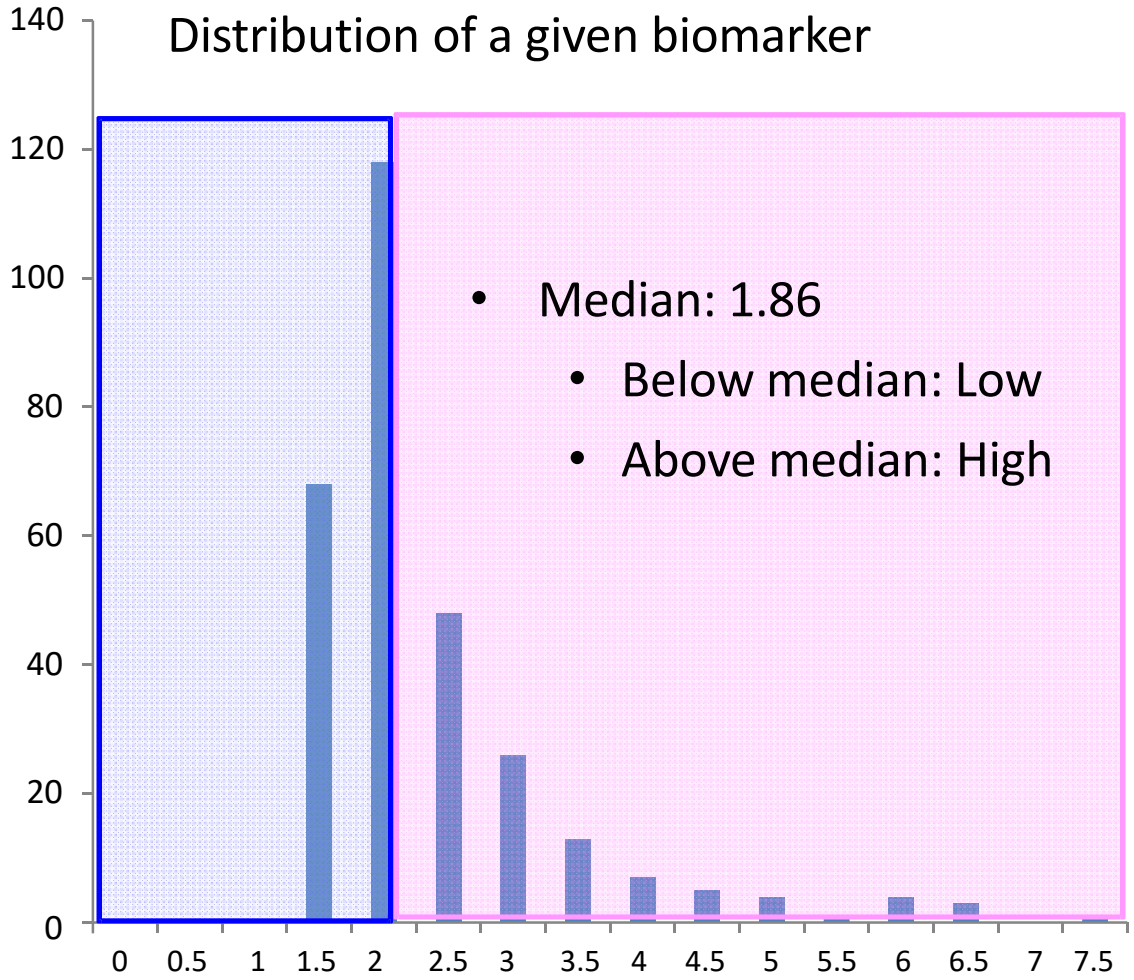
P value for the test of parameter correlation coefficient = 0		Pearson’s correlation coefficient (r)			
		0.1	0.3	0.5	0.7
N	5	0.870	0.615	0.374	0.165
	10	0.783	0.397	0.137	0.022
	20	0.675	0.198	0.024	0.001
	100	0.322	0.002	0.000	0.000
	400	0.045	0.000	0.000	0.000

How to Summarize Two Continuous Variable Datasets

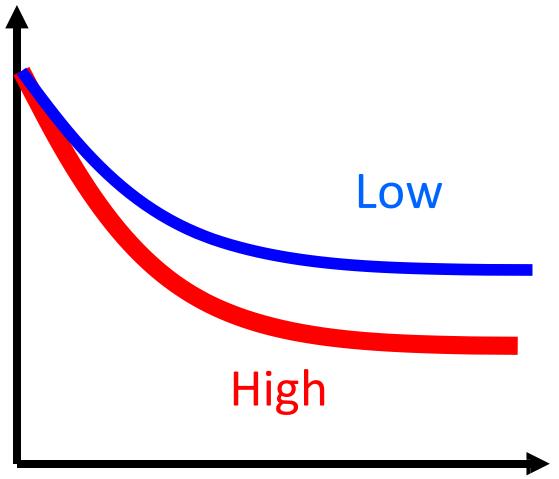
- Use scatter plots
- Check for the presence of outliers and approximate correlations
  - Linear or Curvilinear
- Because clinical data are often distributed with a tail to the right, the Spearman's correlation coefficient is often more appropriate than Pearson's correlation coefficient
- The test results alone do not determine whether there is a correlation.

Do Continuous Variables Have to be Separated by Binary Values?

Examine the relationship between biomarkers and prognosis

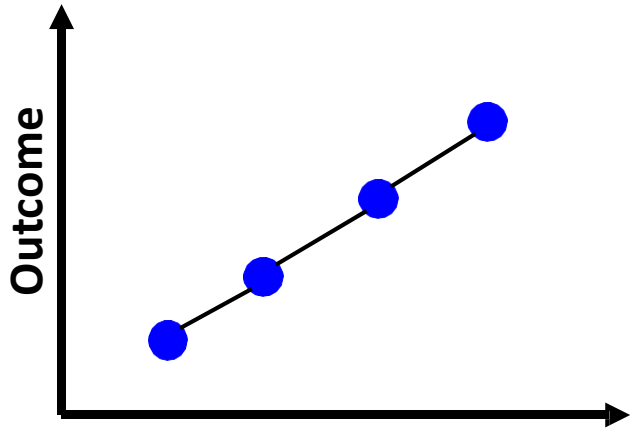


Is this really
what you wanted to find out?

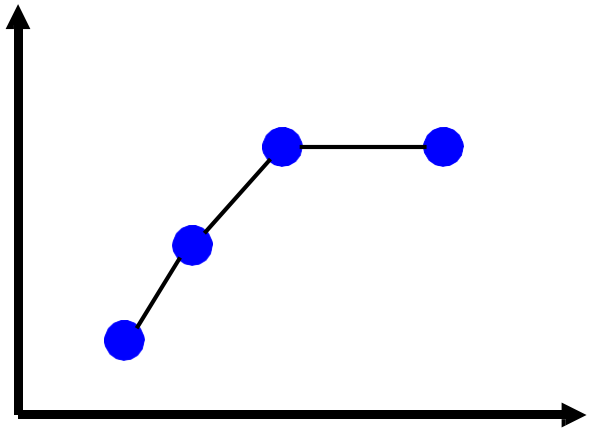


Before Dividing Continuous Variable into Binary Groups

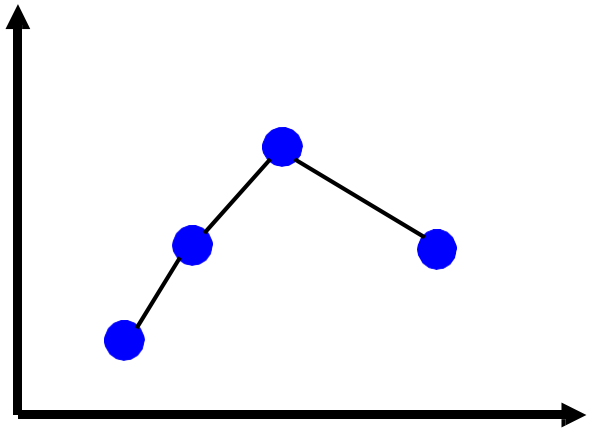
Is there a linear relationship?



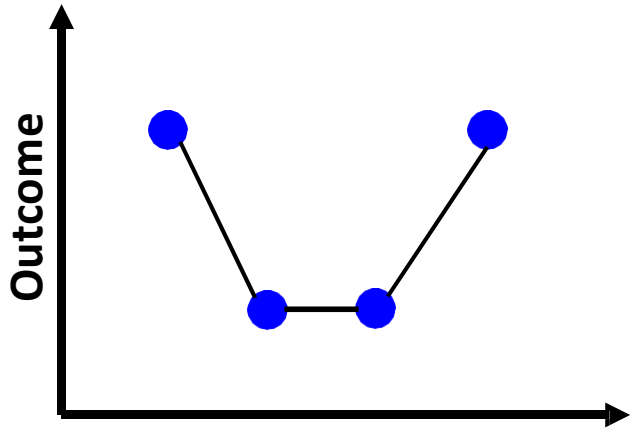
Will the effect plateau?



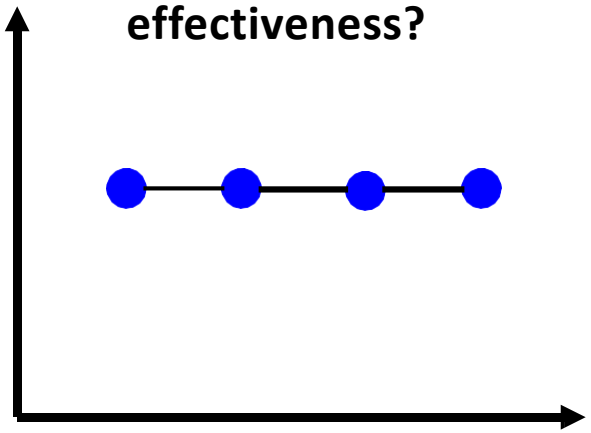
Will the effect be weakened?



Bathtub curve?



No change in effectiveness?



- **When divided into two groups, the nature of the relationship is unclear.**
- **It is more appropriate to consider continuous variables by dividing them into three or more groups.**

How to Summarize Categorical Data

Before performing chi-square test and Fisher’s exact test

Outcome Types and Statistical Methods

	Continuous variable	Binary (0/1)	Survival time
Outcome examples	Blood pressure level Clinical laboratory values	Response rate Adverse events rate	Overall survival Progression-free survival, PFS
Data summarization methods	Histogram Scatter plot	Contingency table	Kaplan–Meier curve
Group comparison (test)	*t*-test Wilcoxon rank sum test	chi-square test Fisher's exact test	Log-rank test
Model fitting	Multiple regression analysis	Logistic regression	Cox regression

Summary Table and Contingency Tables for Categorical Data!

Data set

Patient No.	T factor	N factor	M factor	Stage
1	10	20	10	40
2	10	10	00	10
3	20	20	10	30
4	20	10	00	20
5	30	20	00	30
.	.	.	.	.
.	.	.	.	.

- **Relying solely on Excel's "filter" function is inefficient and carries the risk of overlooking data.**
- **The relationship between the two variables is unclear without checking the contingency table.**

T	Frequency	Percenetage	Cumulative Frequency	Cumulative Percentage
10	2	40.00	2	40.00
20	2	40.00	4	80.00
30	1	20.00	5	100.00

N	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10	1	20.00	1	20.00
10	1	20.00	2	40.00
20	3	60.00	5	100.00

M	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
00	3	60.00	3	60.00
10	2	40.00	5	100.00

stage	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10	1	20.00	1	20.00
20	1	20.00	2	40.00
30	2	40.00	4	80.00
40	1	20.00	5	100.00

Table : M * stage

M / Frequency	10	20	30	40	Sum
00	1	1	1	0	3
10	0	0	1	1	2
Sum	1	1	2	1	5

Contingency Table Analysis

	Group A	Group B
male	XX	XX
female	XX	XX

Chi-square test or Fisher's exact test

Not significant

There is no difference between males and females among groups

Significant

Differences between men and women among groups

Group A has more women than Group B

	Group A	Group B
<40 years old	XX	XX
40–59 years old	XX	XX
>60 years old	XX	XX

Chi-square test or Fisher's exact test

Not significant

No difference between ages among groups

Significant

Differences in age distribution among groups

So what?

Contingency Table Analysis

	No Complications	Complications
<40 years old	15	5
40–59 years old	10	10
>60 years old	5	15

	No Complications	Complications
<40 years old	10	10
40–59 years old	15	5
>60 years old	5	15

With the Fisher's exact test, the *P* value is the same for both (p = 0.0084)

↓

When significant, both interpretations are significantly "different somewhere" (alternative hypothesis)

↓

For "trend test" (Mantel test etc.),

Using JMP, Sample scripts of the Cochran–Armitage test can be downloaded

↓

p = 0.0017: significant
Complications tend to be significantly higher in older people.

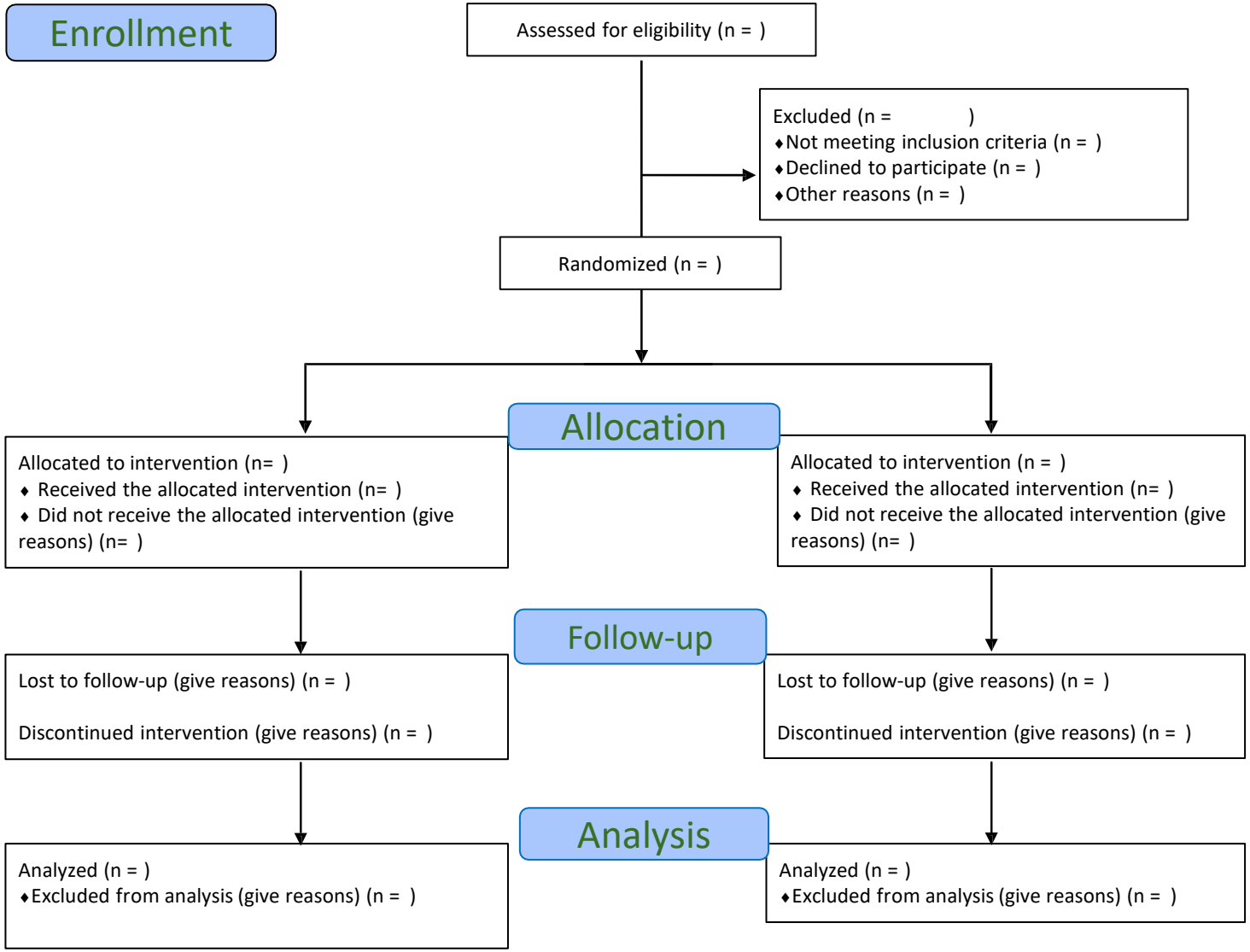
↓

p = 0.1169: Not significant
It cannot be said that complications are more common in older people.

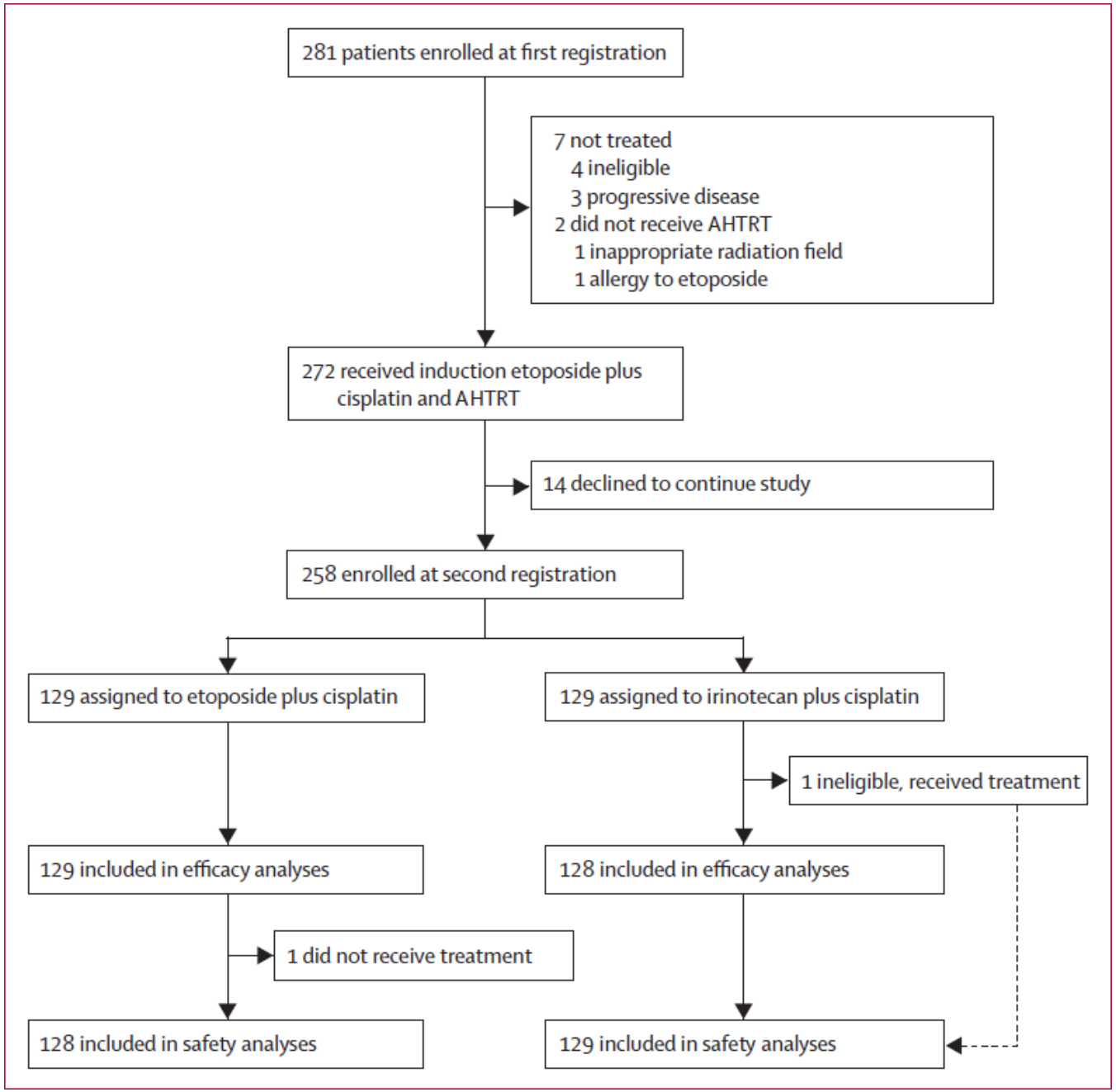
Organize what you want to make sure about before testing!

Analysis of Population and Patient Flow Diagram

RCT: CONSORT Flow Diagram

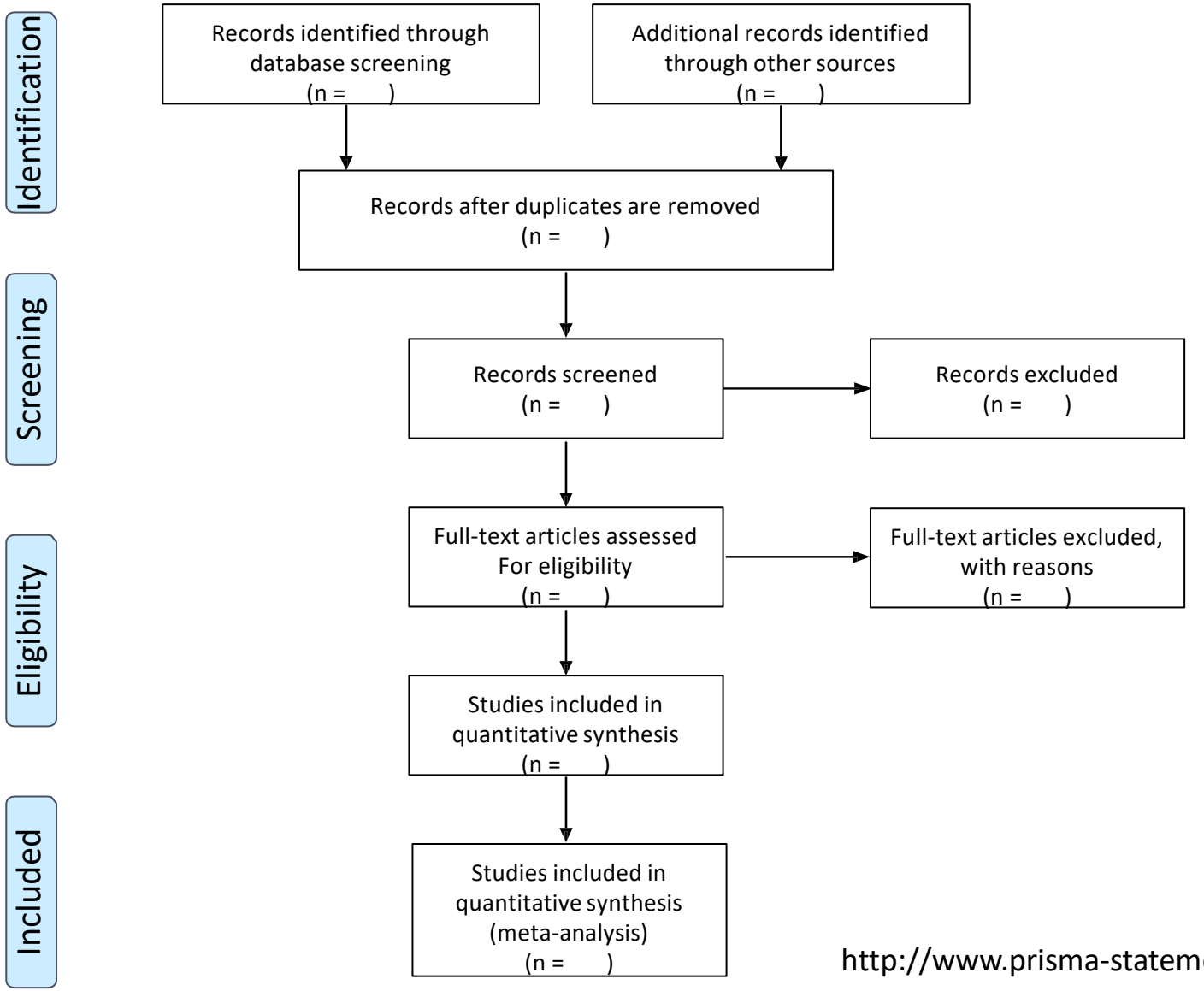


https://pmc.ncbi.nlm.nih.gov/articles/PMC6398298/figure/F1/

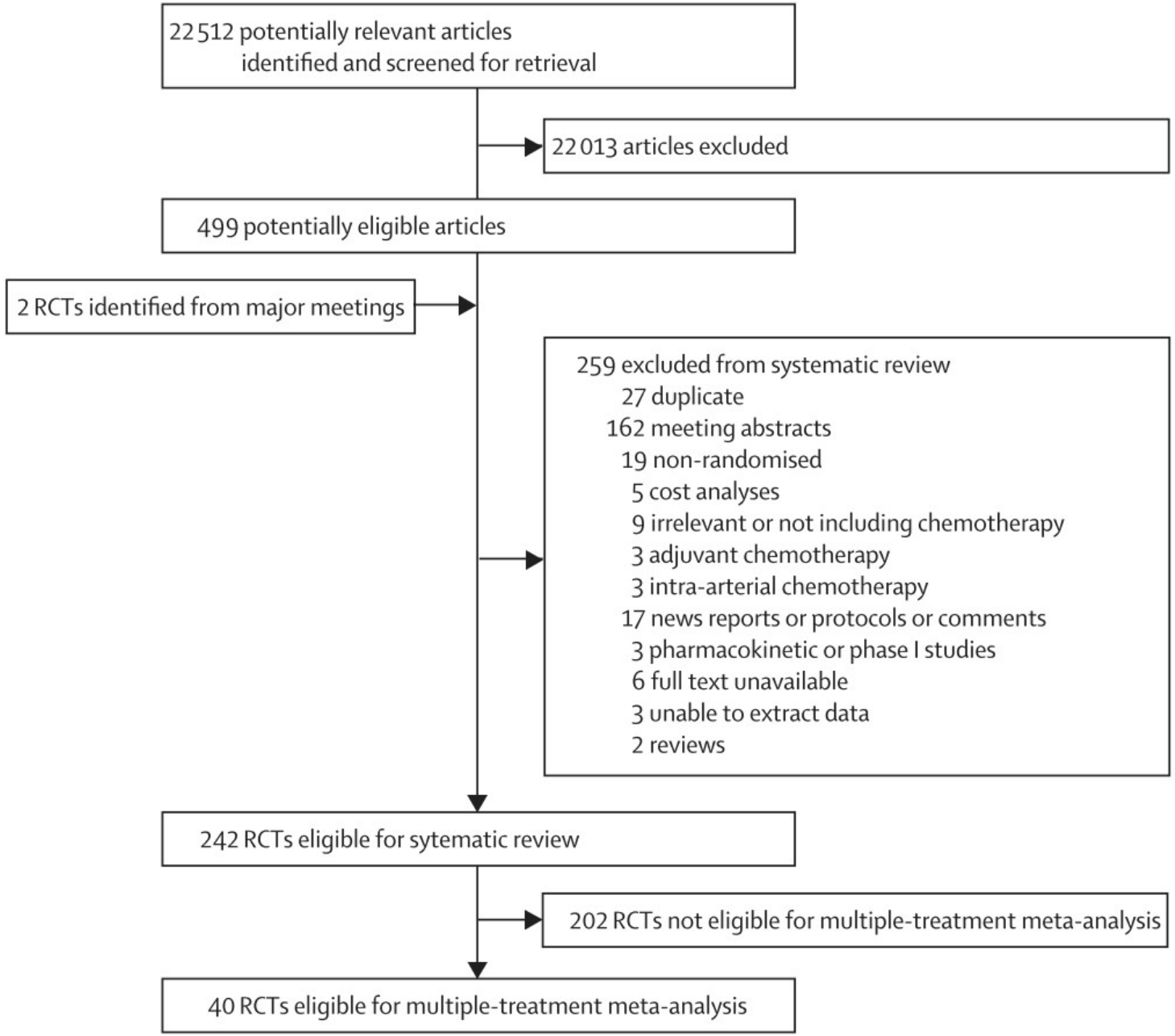


Kubota K, et al. *The Lancet Oncology.* 2014; 15(1): 106-13.

Meta-Analysis: PRISMA Flow Diagram

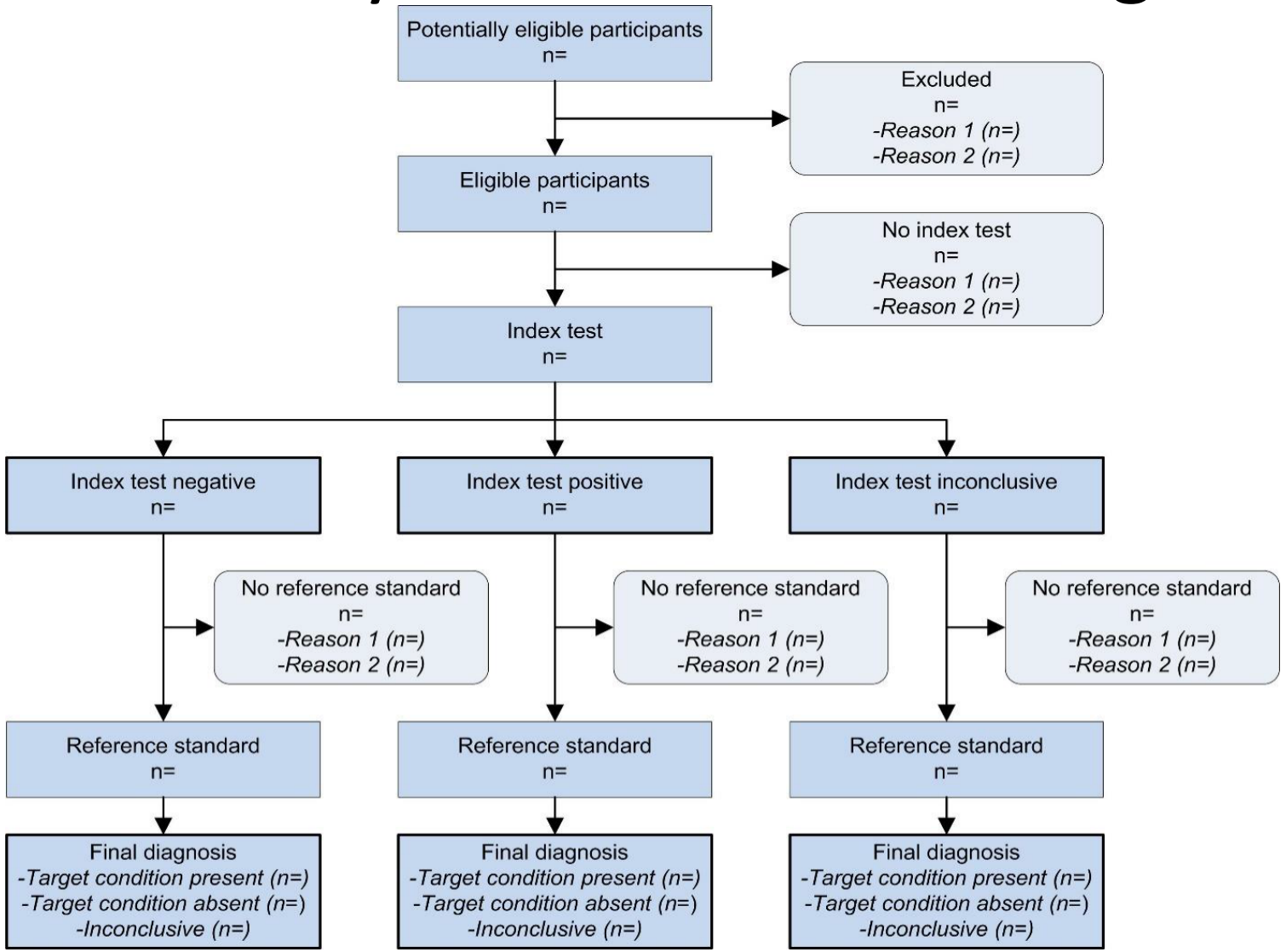


http://www.prisma-statement.org/



Golfinopoulos V et al., *The Lancet Oncology.* 2007; 8(10): 898-911.

Diagnostic Accuracy Studies: STARD Diagram



http://www.equator-network.org/wp-content/uploads/2015/03/STARD-2015-flow-diagram.pdf

Prognostic Factor Analysis

Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK)

Lisa M. McShane, Douglas G. Altman, Willi Sauerbrei, Sheila E. Taube, Massimo Gion, Gary M. Clark for the Statistics Subcommittee of the Working Group on Cancer Diagnostics

Table 1. Reporting recommendations for tumor marker prognostic studies (REMARK)

INTRODUCTION

1. State the marker examined, the study objectives, and any prespecified hypotheses.

MATERIALS AND METHODS

Patients

2. Describe the characteristics (e.g., disease stage or comorbidities) of the study patients, including their source and inclusion and exclusion criteria.
3. Describe treatments received and how chosen (e.g., randomized or rule-based).

Specimen characteristics

4. Describe type of biological material used (including control samples) and methods of preservation and storage.

Assay methods

5. Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.

Study design

6. State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7. Precisely define all clinical endpoints examined.
8. List all candidate variables initially examined or considered for inclusion in models.
9. Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.

Statistical analysis methods

10. Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11. Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.

RESULTS

Data

12. Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.
13. Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including

An

17. Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18. If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.

DISCUSSION

19. Interpret the results in the context of the prespecified hypotheses and other relevant studies; include a discussion of limitations of the study.
20. Discuss implications for future research and clinical value.

Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout.

f research and hundreds of reports on tumor
cology, the number of markers that have
ically useful is pitifully small. Often, initially
s of a marker show great promise, but subse-
n the same or related markers yield inconsis-
is or stand in direct contradiction to the
lts. It is imperative that we attempt to under-
ns that multiple studies of the same marker
conclusions. A variety of methodologic prob-
cited to explain these discrepancies. Unfortu-
mor marker studies have not been reported in
on, and published articles often lack sufficient
llow adequate assessment of the quality of the
eralizability of study results. The development
r the reporting of tumor marker studies was
mendation of the National Cancer Institute–
nisation for Research and Treatment of Can-
TC) First International Meeting on Cancer
2000. As for the successful CONSORT initia-
ized trials and for the STARD statement for
ies, we suggest guidelines to provide relevant
out the study design, preplanned hypotheses,
ecimen characteristics, assay methods, and
sis methods. In addition, the guidelines sug-
sentations of data and important elements to
ssions. The goal of these guidelines is to encour-
t and complete reporting so that the relevant
l be available to others to help them to judge
f the data and understand the context in which
apply. [J Natl Cancer Inst 2005;97:1180–4]

f research and hundreds of reports on tumor
gy, the number of markers that have emerged
l is pitifully small *(1–3)*. Often, initially re-
marker show great promise, but subsequent
e or related markers yield inconsistent conclu-
direct contradiction to the promising results.
at we attempt to understand the reasons that
the same marker lead to differing conclusions.
ems have been cited to explain these discrep-
general methodologic differences, poor study
hat are not standardized or lack reproducibility,
e or misleading statistical analyses that are often
sizes too small to draw meaningful conclusions
mple, in retrospective studies, patient populations
toward patients with available tumor specimens.

Specimen availability may be related to tumor size and patient outcome *(12)*, and the quantity, quality, and preservation method of the specimen may affect feasibility of conducting certain assays. There can also be biases or large variability inherent in the assay results, depending on the particular assay methods used *(13–17)*. Statistical problems are commonplace. These problems include underpowered studies or overly optimistic reporting of effect sizes and significance levels due to multiple testing, subset analyses, and cutpoint optimization *(18)*.

Unfortunately, many tumor marker studies have not been reported in a rigorous fashion, and published articles often lack sufficient information to allow adequate assessment of the quality of the study or the generalizability of study results. Such reporting deficiencies are increasingly being highlighted by systematic reviews of the published literature on particular markers or cancers *(19–25)*.

The development of guidelines for the reporting of tumor marker studies was a major recommendation of the National Cancer Institute–European Organisation for Research and Treatment of Cancer (NCI-EORTC) First International Meeting on Cancer Diagnostics (From Discovery to Clinical Practice: Diagnostic Innovation, Implementation, and Evaluation) that was convened in Nyborg, Denmark, in July 2000. The purpose of the meeting was to discuss issues, accomplishments, and barriers in the field of cancer diagnostics. Poor study design and analysis, assay variability, and inadequate reporting of studies were identified as some of the major barriers to progress in this field. One of the working groups formed at the Nyborg meeting was charged with addressing statistical issues of poor design and analysis and with reporting of tumor marker prognostic studies. The guidelines that we present in this commentary are the product of that committee. The Program for the Assessment of Clinical Cancer Tests (PACCT) Strategy Group of the U.S. NCI has also strongly endorsed this effort (http://www.cancerdiagnosis.nci.nih.gov/assessment).

Affiliations of authors: Biometric Research Branch (LMM) and Cancer Diagnosis Program (SET), National Cancer Institute, Bethesda, MD; Medical Statistics Group, Cancer Research UK, Center for Statistics in Medicine, Wolfson College, Oxford, UK (DGA); Institut fuer Medizinische Biometrie und Medizinische Informatik, Universitaetsklinikum Freiburg, Germany (WS); Centro Regionale Indicatori Biochimici di Tumore, Ospedale Civile, Venezia, Italy (MG); OSI Pharmaceuticals, Inc., Boulder, CO (GMC).

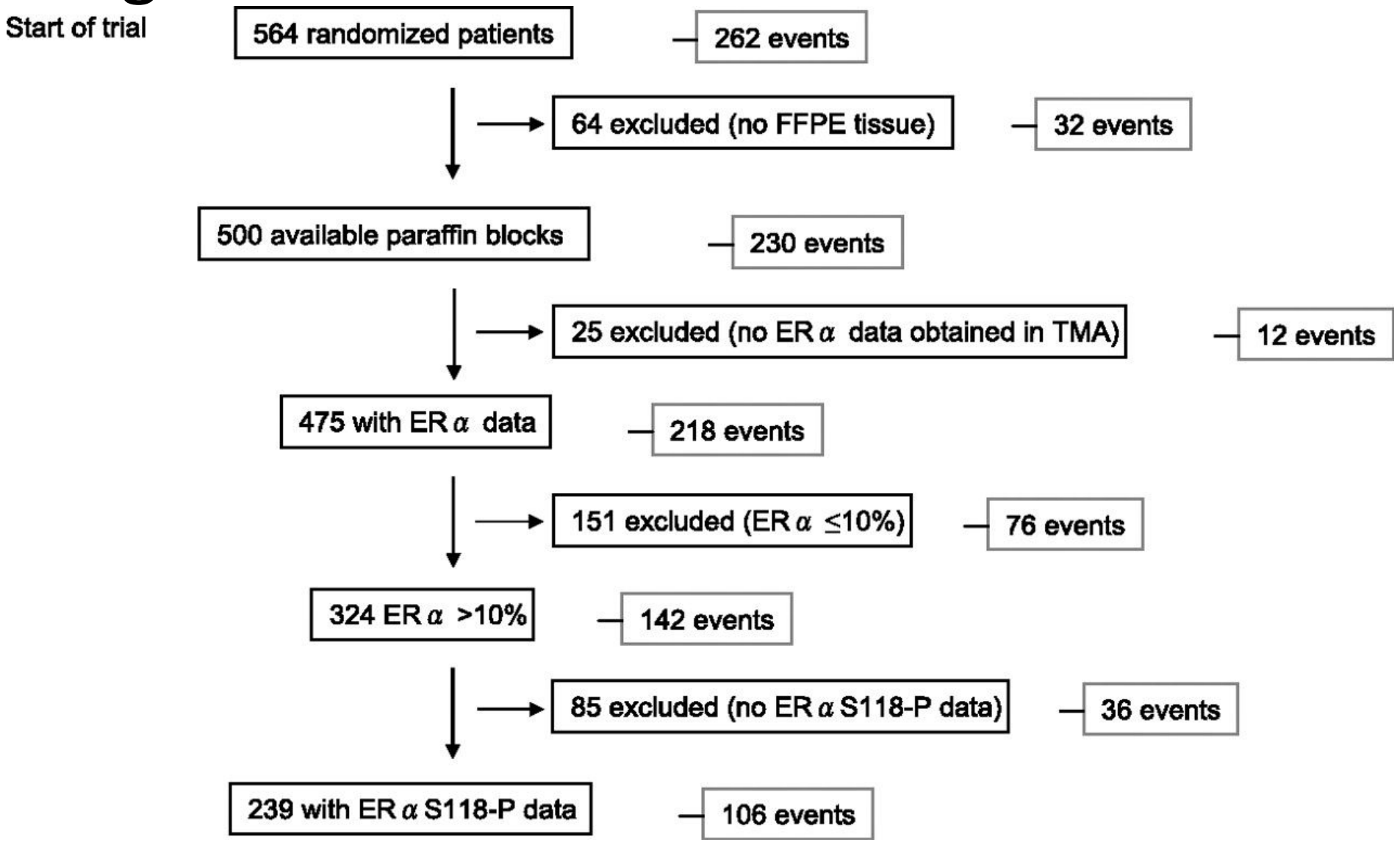
Correspondence to: Lisa M. McShane, PhD, National Cancer Institute, Biometric Research Branch, DCTD, Rm. 8126, Executive Plaza North, MSC 7434, 6130 Executive Blvd., Bethesda, MD 20892-7434 (e-mail: lm5h@nih.gov).

See "Notes" following "References."

DOI: 10.1093/jnci/dji237

McShane LM, Altman DG, Sauerbrei W, et al. *J Natl Cancer Inst* 2005; **97**(16): 1180-4.

Flow Diagram of Patients



Marleen Kok et al. *JNCI J Natl Cancer Inst* 2009; 101: 1725-1729.

The Population to be Analyzed is Best Defined at the Planning Stage

- The subject and data that must be examined in the medical record become clear.
  - What do you want to compare it with?
  - Do the necessary data exist for the comparison?
  - Create flags to distinguish
  - Clarification of research limitation
- If even one subject of analysis changes, all analyses must be redone.

Summary

- Check if the input data are correct before analysis.
  - Double entry or reading review
- Perform a data description rather than testing from the beginning.
  - Check histograms and select appropriate summary statistics and test methods.
- Check the relationship between two variables using scatter plots and contingency tables.
  - Check for outliers and data that should not exist.
- Research with an awareness of the definition of the population to be analyzed.
  - Be aware of which analyses are performed on which subjects.
  - Ensure to draw a patient flow diagram