



Universidad
del Cauca



TYPES OF DATA

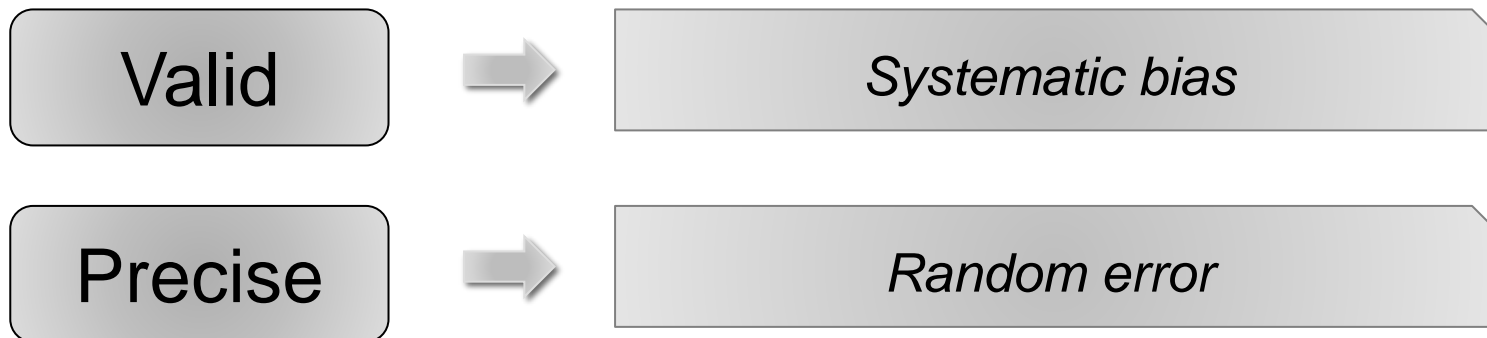
Jose Andres Calvache MD MSc PhD

Department of Anesthesiology, Universidad del Cauca, Colombia

Department of Anesthesiology, Easmus University Medical Centre, Rotterdam, The Netherlands

Basic principles of research

- To obtain a valid and precise estimate of...
 - The efficacy of a certain treatment
 - The relationship between a potential risk factor and an outcome
 - The prevalence of any disease of health related
 - The prognostic value of a single or multiple predictors in ...



Basic principles in data analysis

- Do a lot of graphs
- Use the common sense
- “Critical thinking”
- This is the only beginning
- Look for appropriate advice and support from the start

Basic definitions

Population



Sample



Variable



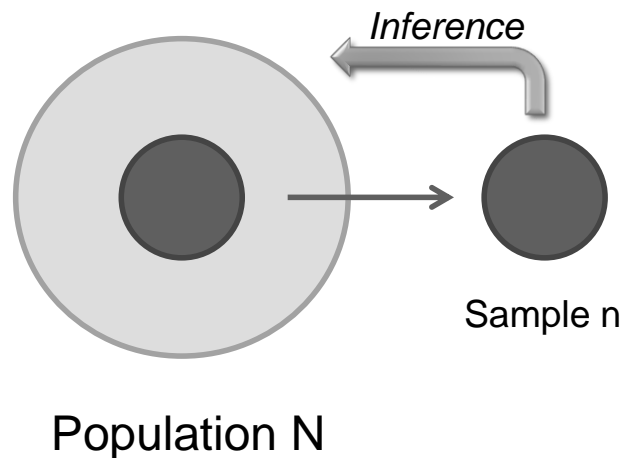
Data

Population

- Group of subjects that share at least **one characteristic**
 - N = Size of the population
 - Student of the university
 - Health care workers in Popayan
 - Public hospitals in Colombia
 - Patients with histologically confirmed diagnosis of gastric cancer
 - Older patients under abdominal surgery (certain age)
- *Population should be very clearly defined*
- *Unit of analysis: member of the population*

Sample

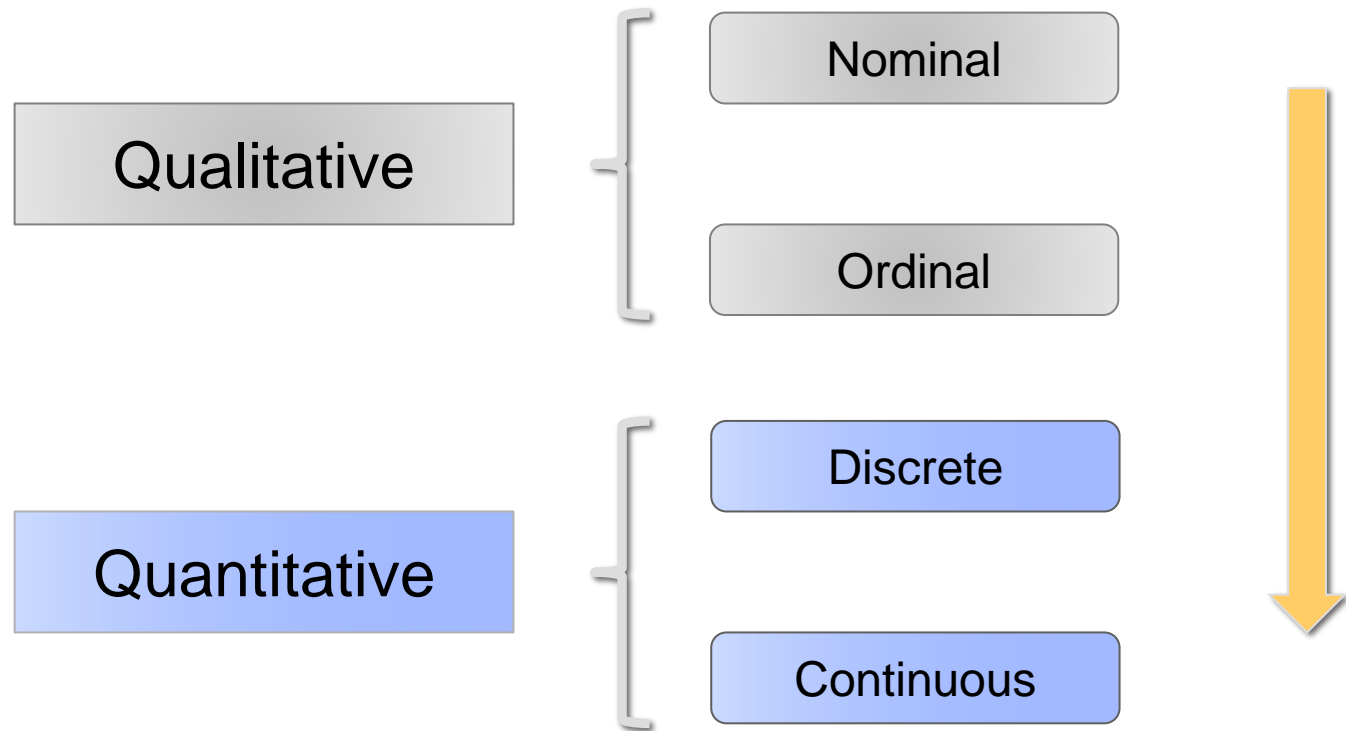
- Subgroup of the population of interest
 - n : sample size
 - “Maximize representativeness of the population”
 - Random sample ?
 - Should be described in detail



Variable

- Characteristic of the population in which we are interested in
 - X_n : univariable
 - X_1, X_2, X_3, X_p : p multivariable
- Name
- Gender, socioeconomical status
- Age, weight, BMI
- Outcomes (fail/success), (death/alive)
- Length of hospital stay
- Should be clearly defined in the protocol (age groups, fail ?)

Variable



Nominal variables

- To express a quality,
 - By using k categories -mutually exclusive- (C_1, C_2, C_n)
 - Also called categorical variables
- Racial group : African, Caucasian, Hispanic, Asian ($k = 4$)
- Blood group : A, B, AB, O ($k = 4$)
- Gender : Male, Female ($k = 2$)
- The number of categories should be clearly defined in advance
- Also called categorical variables (binary $k = 2$) **
- It is possible to transform the variable in $(k - 1)$ or reference category
- Order is meaningless
- How to summarize : Frequency table (n, %)

Ordinal variables

- To express an ordered quality ($C_1 < C_2 < C_n$)
 - Pain level: (None, slight, moderate, severe)
 - Cancer stage : (I, II, III, IV)
 - Social stratum : 1, 2, 3, 4 (numbers as categories)
- Categories could be numbered : (1, 2, 3, n)
- How to summarize : Frequency table (n, %)

How to summarize : Frequency table (n, %)

Pain level after surgery	Freq (n)	%
None	10	20
Slight	15	30
Moderate	15	30
Severe	10	20

Qualitative variables

- Mutually exclusive
 - Could take one or more categories (C_1, C_2, \dots, C_k)
 - Affected systems: (cardiovascular, renal, nervous) ($k = 3$)
 - Risk factors: (smoke, sedentary, hypertension) ($k = 3$)
- Reordering
 - Could be transformed into a binary variable for each category
 - $X_1 = 1$, if C_1 is present. Otherwise 0
 - $X_2 = 1$, if C_2 is present. Otherwise 0
 - $\dots X_n$

Discrete -quantitative- variables

- To express a numerical quantity
- Usually calculated by counting or measurements
- **Discrete variables (counting)**
 - They take a finite number of values (k) without decimals
 - Number of childrens at school
 - Number of medical consultations

Continuous -quantitative- variables

- All possible real values in an interval or in an continuum
 - Usually associated with an unit of measurement
 - Weigth (kg)
 - BMI (kg/m^2)
 - Cholesterol level (mmol/L)
 - Survival time in certain cancer (years, months)
 - Systolic blood pressure (mmHg)
- How to summarize these variables ?

Data

- A single observation of a variable in the population
- Data will be analyzed
- Single value for BMI
- Single value for weight
- Single value for age

Data sample

- n observations in n subjects (rows versus columns)

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

Subject	Age (years)	Weight (kg)	Height (cm)	BMI (kg/m ²)	Heart rate (bpm)	Glucose (g/L)
1	50	75	182	22.6	111	0.84
2	21	76	178	24.0	138	1.04
3	42	76	175	24.8	110	0.51
4	53	71	168	25.2	103	0.81
5	42	76	187	21.7	117	0.63
6	45	69	171	23.6	122	0.63
7	30	76	176	24.5	124	0.87
8	21	60	183	17.9	124	0.67
9	19	63	173	21.0	127	0.80
10	17	64	175	21.6	131	0.95

Research

Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study

BMJ 2020 ; 368 doi: <https://doi.org/10.1136/bmj.l6669> (Published 08 January 2020)

Cite this as: BMJ 2020;368:l6669

[Article](#)[Related content](#)[Metrics](#)[Responses](#)[Peer review](#)

Yanping Li , senior research scientist ¹, Josje Schoufour, senior researcher ^{2 3}, Dong D Wang, research scientist ¹, Klodian Dhana, assistant professor ^{1 4}, An Pan, professor ⁵, Xiaoran Liu, research associate ¹, Mingyang Song, assistant professor ^{1 6 7 8}, Gang Liu, professor ^{1 9}, Hyun Joon Shin, instructor in medicine ¹⁰, Qi Sun, associate professor ^{1 11}, Laila Al-Shaar, postdoctoral research fellow ¹, Molin Wang, associate professor ⁶, Eric B Rimm, professor ^{1 11}, Ellen Hertzmark, professor ¹², Meir J Stampfer, professor ^{1 6 11}, Walter C Willett, professor ^{1 6 11}, Oscar H Franco, professor ^{2 13}, Frank B Hu, professor ^{1 6 11}

Objective: To examine how a healthy lifestyle is related to life expectancy that is free from major chronic diseases.

Design: Prospective cohort study.

Setting and participants: The Nurses' Health Study (1980-2014; n=73 196) and the Health Professionals Follow-Up Study (1986-2014; n=38 366).

Main exposures: Five low risk lifestyle factors: never smoking, body mass index 18.5-24.9, moderate to vigorous physical activity (≥ 30 minutes/day), moderate alcohol intake (women: 5-15 g/day; men 5-30 g/day), and a higher diet quality score (upper 40%).

Main outcome: Life expectancy free of diabetes, cardiovascular diseases, and cancer.

Results: The life expectancy free of diabetes, cardiovascular diseases, and cancer at age 50 was 23.7 years (95% confidence interval 22.6 to 24.7) for women who adopted no low risk lifestyle factors, in contrast to 34.4 years (33.1 to 35.5) for women who adopted four or five low risk factors. At age 50, the life expectancy free of any of these chronic diseases was 23.5 (22.3 to 24.7) years among men who adopted no low risk lifestyle factors and 31.1 (29.5 to 32.5) years in men who adopted four or five low risk lifestyle factors. For current male smokers who smoked heavily (≥ 15 cigarettes/day) or obese men and women (body mass index ≥ 30), their disease-free life expectancies accounted for the lowest proportion ($\leq 75\%$) of total life expectancy at age 50.

Conclusion: Adherence to a healthy lifestyle at mid-life is associated with a longer life expectancy free of major chronic diseases.

Table 1 | Participants' characteristics in middle of follow-up period according number of low risk lifestyle factors in Nurses' Health Study (women) and Health Professionals Follow-up Study (men). Values are numbers (percentages) unless stated otherwise*

	No of low risk lifestyle factorst				
	Zero	One	Two	Three	Four or five
Nurses' Health Study (1998)	(n=11 749)	(n=25 048)	(n=19 837)	(n=9079)	(n=2984)
Mean (SD) age, years	63.3 (7.1)	63.6 (7.2)	63.7 (7.2)	63.8 (7.1)	63.6 (6.9)
Mean (SD) body mass index	31.0 (5.3)	29.2 (5.5)	26.6 (5.0)	24.4 (3.5)	23.1 (2.1)
Mean (SD) Alternate Healthy Eating Index score	43.0 (6.8)	46.6 (9.2)	51.4 (9.8)	55.3 (9.3)	58.3 (8.1)
Mean (SD) physical activity, h/week	0.5 (1.0)	0.9 (1.8)	1.9 (3.0)	3.5 (3.8)	5.5 (4.3)
Mean (SD) alcohol consumption, g/day	5.0 (10.8)	4.5 (9.1)	5.1 (8.1)	6.1 (7.4)	7.3 (5.5)
Past smoking	9086 (77.5)	11 314 (45.2)	6822 (34.4)	2658 (29.3)	558 (18.6)
Current smoking	2663 (22.5)	3249 (13.0)	1654 (8.4)	409 (4.5)	50 (1.7)
White	11 524 (98.1)	24 496 (97.8)	19 289 (97.2)	8839 (97.4)	2921 (97.9)
Multivitamin use	5412 (46.1)	12 437 (49.6)	10 480 (52.8)	5220 (57.5)	1765 (59.1)
Regular aspirin use	5212 (44.5)	11 457 (45.7)	9145 (46.1)	4358 (47.9)	1535 (51.6)
Family history of diabetes	3583 (30.5)	7394 (29.5)	5296 (26.7)	2170 (23.9)	690 (23.1)
Family history of cancer	1549 (13.3)	3438 (13.7)	2721 (13.7)	1236 (13.6)	419 (14.1)
Family history of myocardial infarction	3153 (26.8)	6273 (25.1)	4787 (24.1)	2069 (22.8)	705 (23.6)