Type of variables

Age is a variable whose possible values are non-negative.

Gender is a different sort of variable – its possible values (male, female) are not numbers, but types, or categories.

Cause of death is the same kind of variable as gender – its possible values are also categories (heart disease, cancer, suicide, etc.)

Variables like age and number of births, whose possible values are numbers, are called **numerical** variables (naturally). Variables like gender and cause of death are called **categorical** variables.

Both numerical and categorical variables are further classified according to subtypes:

Categorical variables:

Nominal (unordered)	male / female
	smoker / nonsmoker
Ordinal (ordered)	nonsmoker / light / heavy

Numerical variables:

Discrete	number of births
	number of cigarettes
	distance (to nearest mile)
	height (to nearest inch)
Continuous	blood pressure
	temperature
	Height
	distance

Sometimes we use numbers to label categories,

e.g., male=1, female=2

This does not change a categorical variable (gender, in this instance) into a numerical one. The numbers are simply being used as symbols, just like β and θ. It does not make sense to treat them like numbers …

female = twice male ?

male = 1/2 female ?

male < female ?

Types of Studies

Experimental:	Randomized clinical trials
	Laboratory experiments
	Others…
Observational:	Retrospective reviews
	Case-control studies
	Cohort studies
	Surveys
	Others…

The difference:

Experimental studies are often designed to control some variables connected with the experiment. (Observational studies often try to do this in the data analysis stage, where it *may* be too late.)

Experimental and observational studies have different strengths and weaknesses.

Both provide data that can be evaluated statistically. Yet experimental data is often considered better or stronger than observational data. This is a consequence of the context from which the data comes (goes to generalizability) and not the data themselves!

- We can learn from observational studies.

Longitudinal versus Cross-sectional studies

Cross-sectional studies are investigations at a single point in time.

Longitudinal studies are investigations over a period of time.

Context affects inference / generalizability:

- Both can be either an Experimental or Observational study
- A Cross-sectional study measures association at a single point in time.
- A Longitudinal study measures, in addition to association, causation. (Although how to do this is still being debated.)

Issues with context

- Is the sample representative of the target population?
- Does the experimental setting mimic the clinical setting? ('Hawthorne effect')
- Are groups comparable in a comparative study?
- Precisions and validity of measurements?
- Properly constructed instruments for gathering data?

The key to a well-designed study is having an attainable, well-defined (often means narrow) question of interest.

Women's Health Initiative: Overall Risks and Benefits

Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women

Principal Results From the Women's Health Initiative Randomized Controlled Trial.
JAMA 2002;288:321-333

After a mean of 5.2 years of follow-up, the trial was stopped because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits.

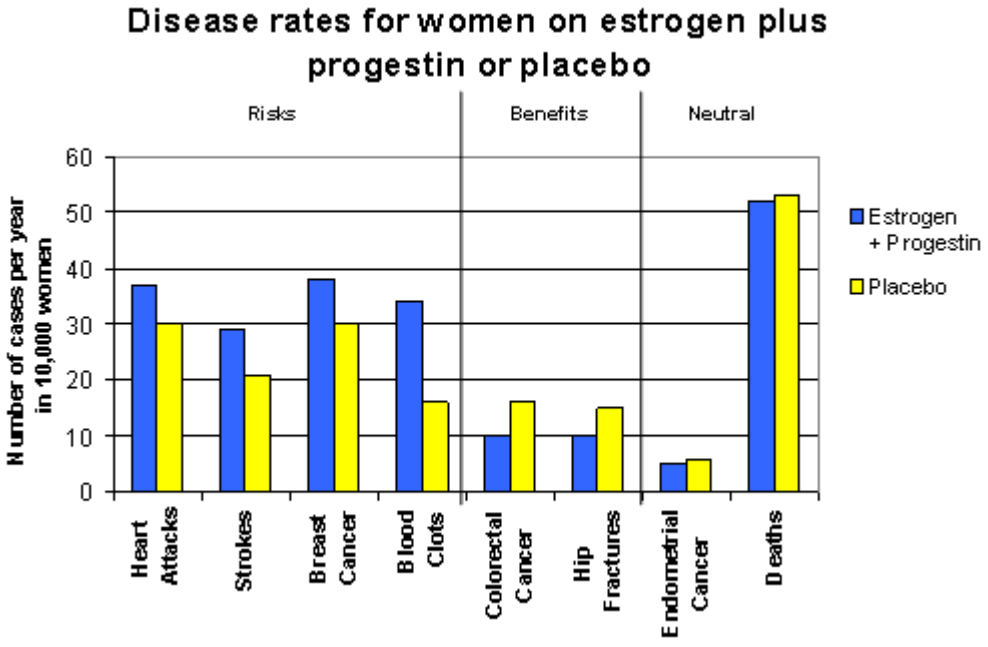
(We'll soon learn what this means technically)

"The risk-benefit profile found in this trial is not consistent with the requirements for a viable intervention for primary prevention of chronic diseases, and the results indicate that this regimen should not be initiated or continued for primary prevention of CHD. "

Abstract: http://jama.ama-assn.org/cgi/content/abstract/288/3/321

Results: Hazard ratios (nominal 95% confidence intervals) were as follows:

CHD: 1.29 (1.02-1.63) with 286 cases
Breast cancer: 1.26 (1.00-1.59) with 290 cases;
Stroke: 1.41 (1.07-1.85) with 212 cases;
Pulmonary embolism: 2.13 (1.39-3.25) with 101 cases
Colorectal cancer: 0.63 (0.43-0.92) with 112 cases
Endometrial cancer: 0.83 (0.47-1.47) with 47 cases
Hip fracture: 0.66 (0.45-0.98) with 106 cases
Death (other causes): 0.92 (0.74-1.14) with 331 cases
Total mortality: 0.98 (0.82-1.18)
Global index: 1.15 (1.03-1.28)



Absolute excess risks per 10,000 person-years attributable to estrogen plus progestin were 7 more CHD events, 8 more strokes, 8 more pulmonary emboli, and 8 more invasive breast cancers, while absolute risk reductions per 10,000 person-years were 6 fewer colorectal cancers and 5 fewer hip fractures. The absolute excess risk of events included in the global index was 19 per 10,000 person-years.

"Overall health risks exceeded benefits from use of combined estrogen plus progestin for an *average* 5.2-year follow-up among healthy postmenopausal US women. All-cause mortality was not affected during the trial."

Q: How are the intervals calculated?

This question is important because different methods lead to different intervals and in this case a common adjustment to the interval was ignored (notice the hint: nominal) which resulted in significantly shorter intervals. The properly adjusted intervals are much wider and cover the HR of 1 in all cases. (See primary JAMA paper)

Lesson: The result is always dependent on the method, so interpreting the result means interpreting the methods as well.