Applied Biostatistics

https://moodle.epfl.ch/course/view.php?id=15590

- Clinical trials intro : phases
- Biostatistical aspects of study protocol
- Study designs
- Statistical analyses
- Power/sample size analysis
- Simulation studies

Clinical trials intro : phases

Summary of clinical trial phases						
Phase	**Primary goal**	**Dose**	**Patient monitor**	**Typical number of participants**	**Success rate[2]**	**Notes**
Preclinical	Testing of drug in non-human subjects, to gather efficacy, toxicity and pharmacokinetic information	unrestricted	scientific researcher	not applicable (*in vitro* and *in vivo* only)		
Phase 0	Pharmacokinetics; particularly, oral bioavailability and half-life of the drug	very small, subtherapeutic	clinical researcher	10 people		often skipped for phase I
Phase I	Testing of drug on healthy volunteers for dose-ranging	often subtherapeutic, but with ascending doses	clinical researcher	20–100 normal healthy volunteers (or for cancer drugs, cancer patients)	approximately 70%	determines whether drug is safe to check for efficacy
Phase II	Testing of drug on patients to assess efficacy and side effects	therapeutic dose	clinical researcher	100–300 patients with specific diseases	approximately 33%	determines whether drug can have any efficacy; at this point, the drug is not presumed to have any therapeutic effect whatsoever
Phase III	Testing of drug on patients to assess efficacy, effectiveness and safety	therapeutic dose	clinical researcher and personal physician	300–3,000 patients with specific diseases	25–30%	determines a drug's therapeutic effect; at this point, the drug is presumed to have some effect
Phase IV	Postmarketing surveillance – watching drug use in public	therapeutic dose	personal physician	anyone seeking treatment from their physician	N/A	watch drug's long-term effects

Biostatistical aspects

- Background/rationale, justification for current study
- *Specific* objective(s)/research question(s) ; consider :
  - The objective of this investigation is to assess the efficacy of drug D in hypertensive patients, *vs.*
  - The objective of this investigation is to assess whether drug D is superior to placebo P in the treatment of hypertensive patients with diastolic blood pressure (DBP) between 90 and 105 mm Hg for six months
- If more than one objective, which is primary *vs.* secondary
- Study plan
  - enroll and treat patients, monitor the study, ensure patient safety, collect valid data
  - describe procedures to be used in the diagnoses, treatment, management of patients
- Study Population

Biostatistical aspects : study design I

- Type of study :
  - Is the study prospective ?
  - Control type (placebo, positive, historical, etc.) ?
  - Single or multi-center ?
  - Parallel study, crossover, stratified, some other type ?
- Treatment group specification and assignment
  - specify treatment groups and interventions (drug, dose, etc.) that patients in the groups will receive
  - how will patients be assigned to the treatment groups to remove assignment bias
  - *gold standard :* randomly assign patients to the groups in *balanced fashion*
  - (Minor departures from balance might sometimes be preferable, for example assigned twice as many subjects to the treatment as to the placebo ; this 2-to-1 balance departure should have small impact on power)

Biostatistical aspects : study design II

- Guaranteeing blinding
- Concomitant medications/treatments
- All protocol procedures :
  - enrolling, diagnosing, treating, or medically monitoring patients
  - applies to all protocol phases : applies to all phases : pre-treatment, during treatment, or post treatment

Biostatistical aspects, cont.

- Guaranteeing blinding
- Concomitant medications/treatments
- All protocol procedures :
  - enrolling, diagnosing, treating, or medically monitoring patients
  - applies to all protocol phases : applies to all phases : pre-treatment, during treatment, or post treatment
- Problem management : define criteria for dealing with problems that could arise, such as
  - significant changes in clinical laboratory parameters
  - severe adverse events
  - actions to be taken for protocol deviations or violations

Biostatistical aspects : statistical analysis

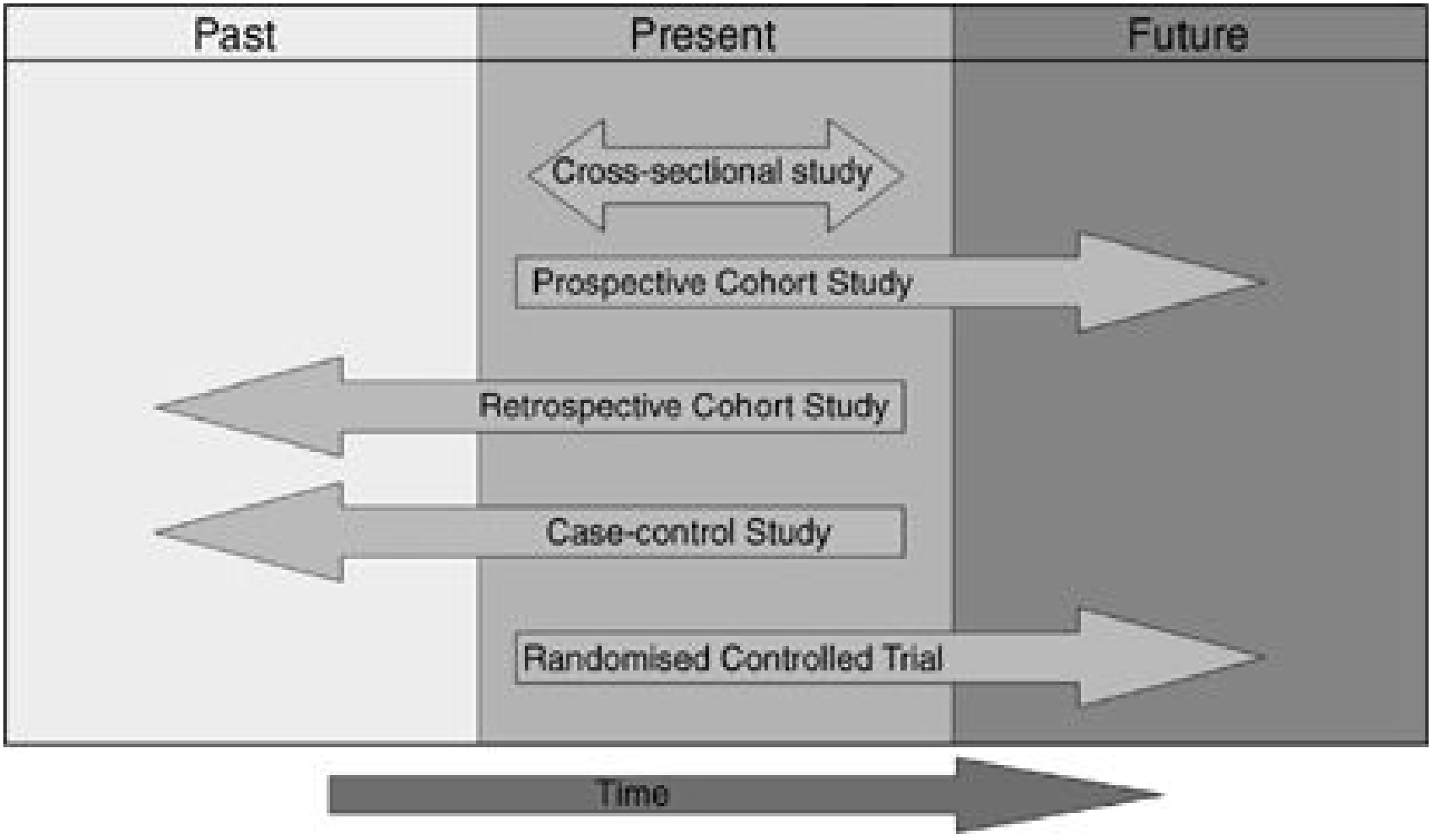
- Formulating objectives as statistical hypotheses
- What endpoints are to be analyzed
  - dichotomous
  - categorical (nominal/ordinal)
  - quantitative
  - (censored) survival time
- Analysis methods
  - logistic regression
  - χ^2 testing
  - general linear model (regression/anova/ancova)
  - survival methods : Kaplan-Meier, Cox regression, *etc.*
- Statistical monitoring procedures
  - sample sizes for early termination
  - group sequential procedures
- Subset analysis

sometimes a trial is supposed to last a certain amount of time, but sometimes you find that the treatment is so much better (or worse) than thte placebo that you decide to stop the trial.

Study designs

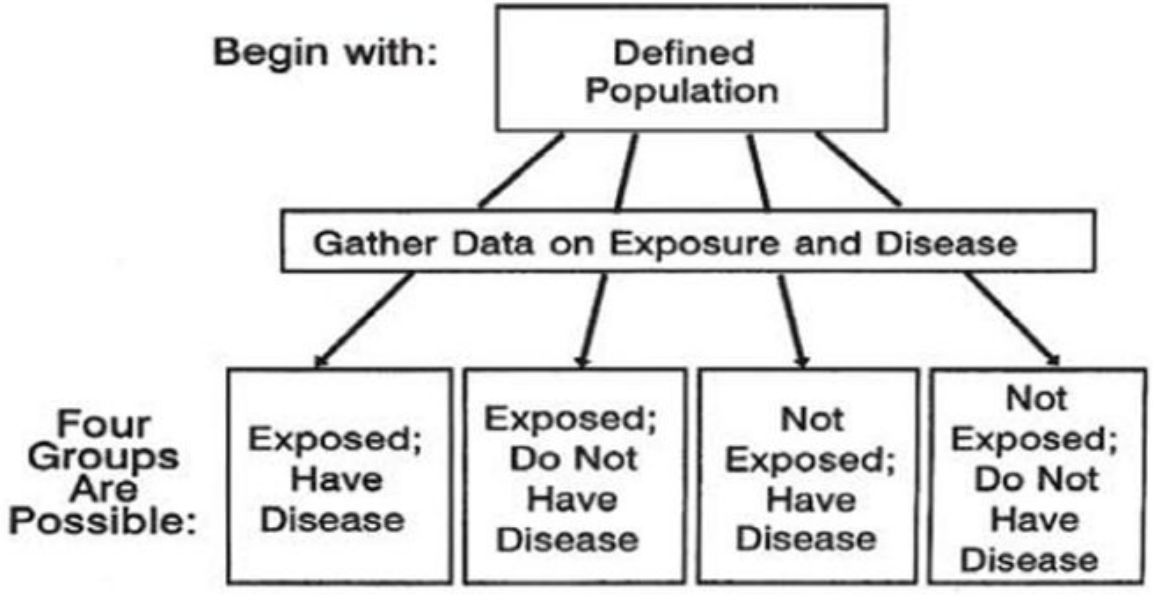
- Design can be considered more important than the analysis : a badly designed study can never be retrieved, whereas a poorly analysed one can usually be reanalysed
- Consideration of design is also important because the design of a study will govern how the data are to be analysed
- Most medical studies consider an *input* (*e.g.* an intervention) and an *output* (*e.g.* some measure of health) T
- One way to categorise studies is with reference to the time sequence in which the input and output are studied

Time sequence

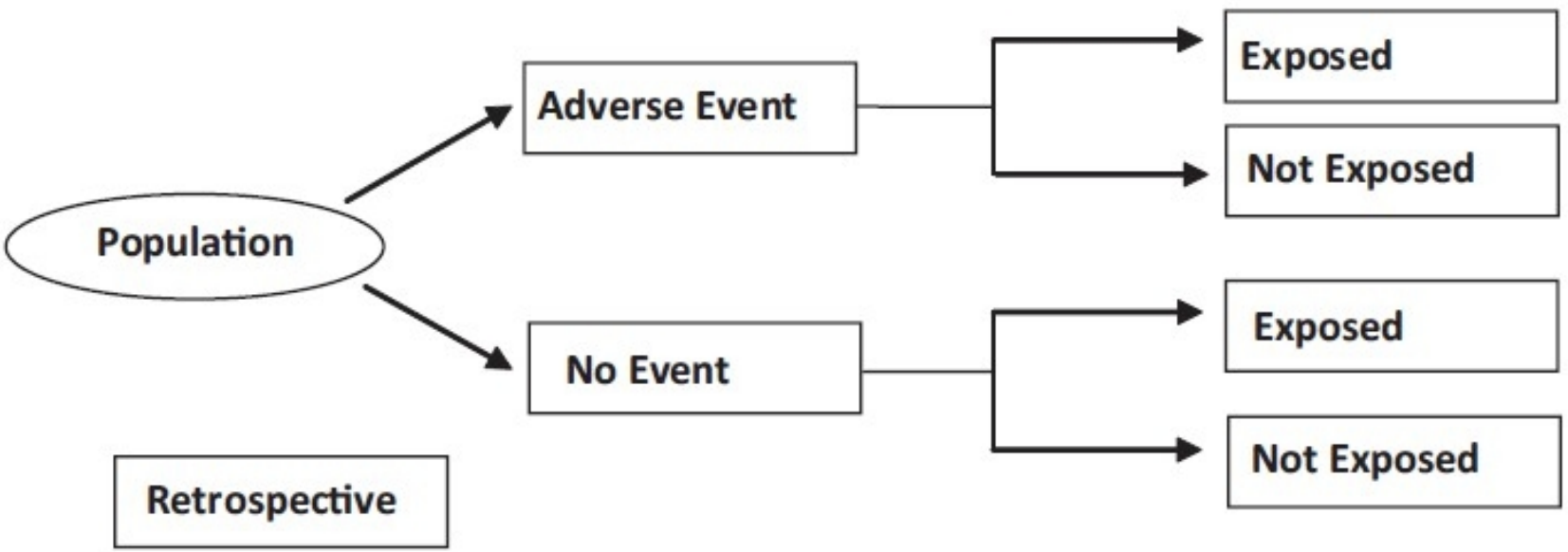


Cross-sectional study

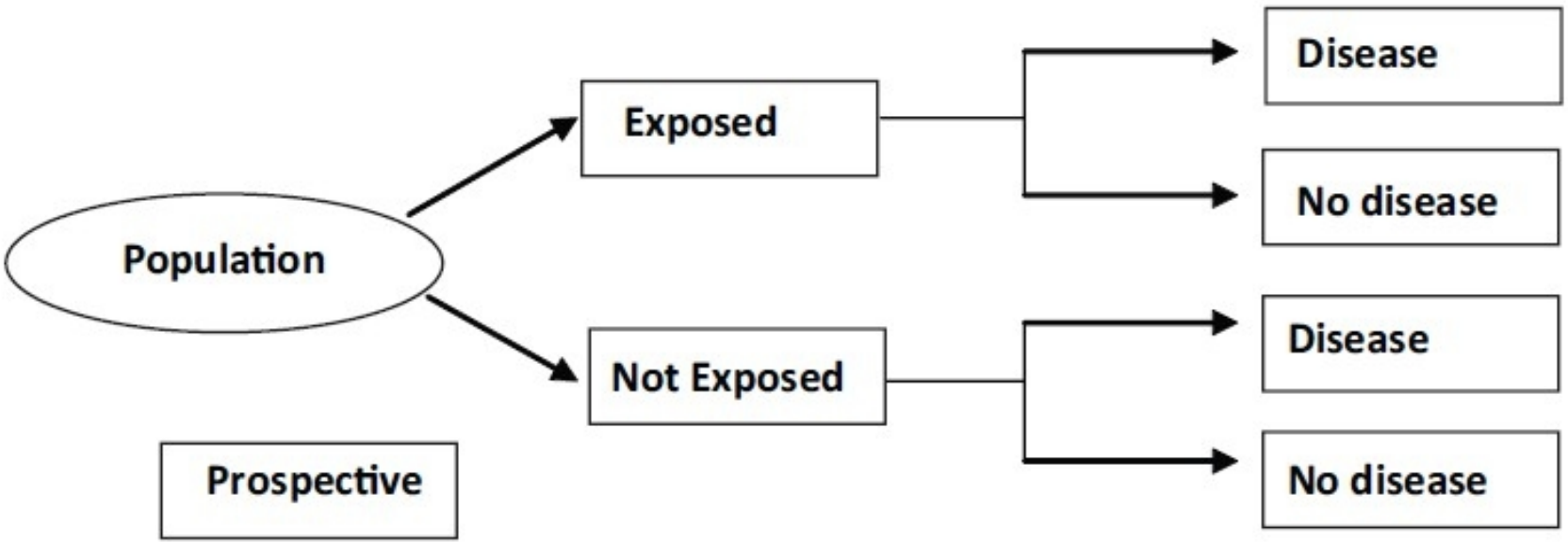
Design of cross sectional study



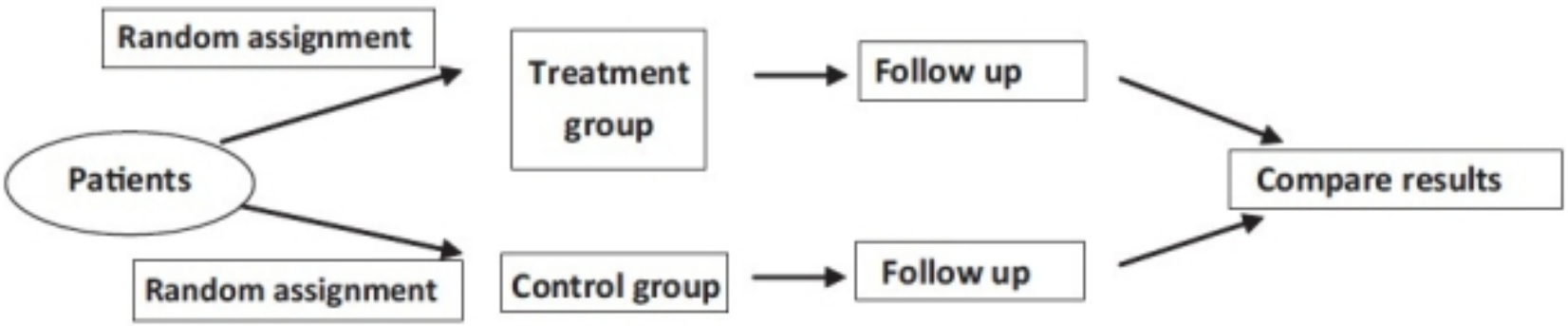
Case-control study



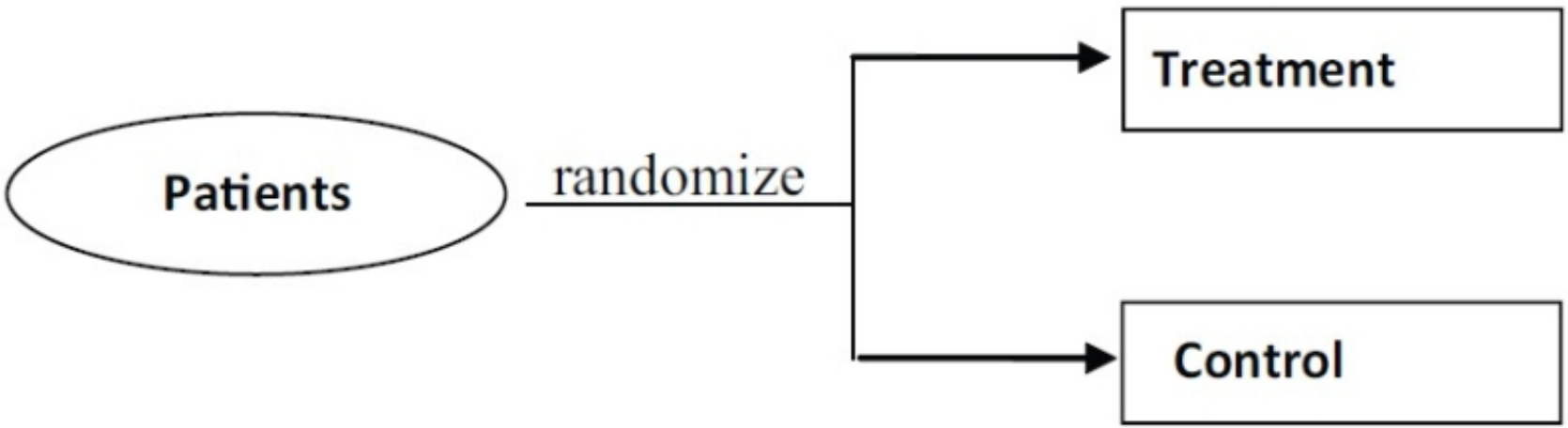
Cohort (longitudinal) study



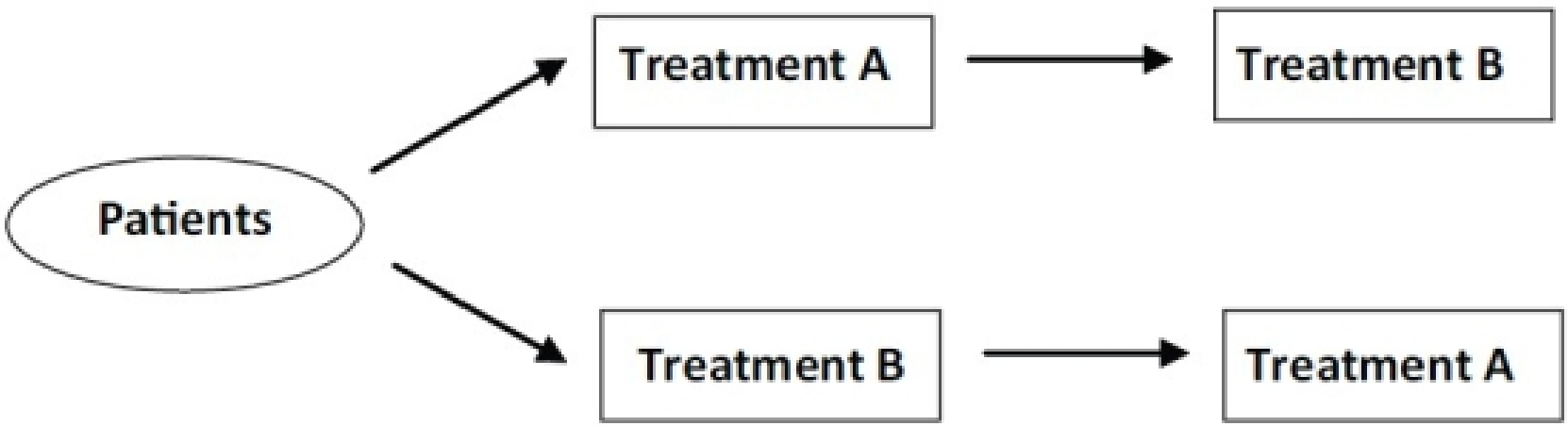
Randomized clinical trial



Parallel design

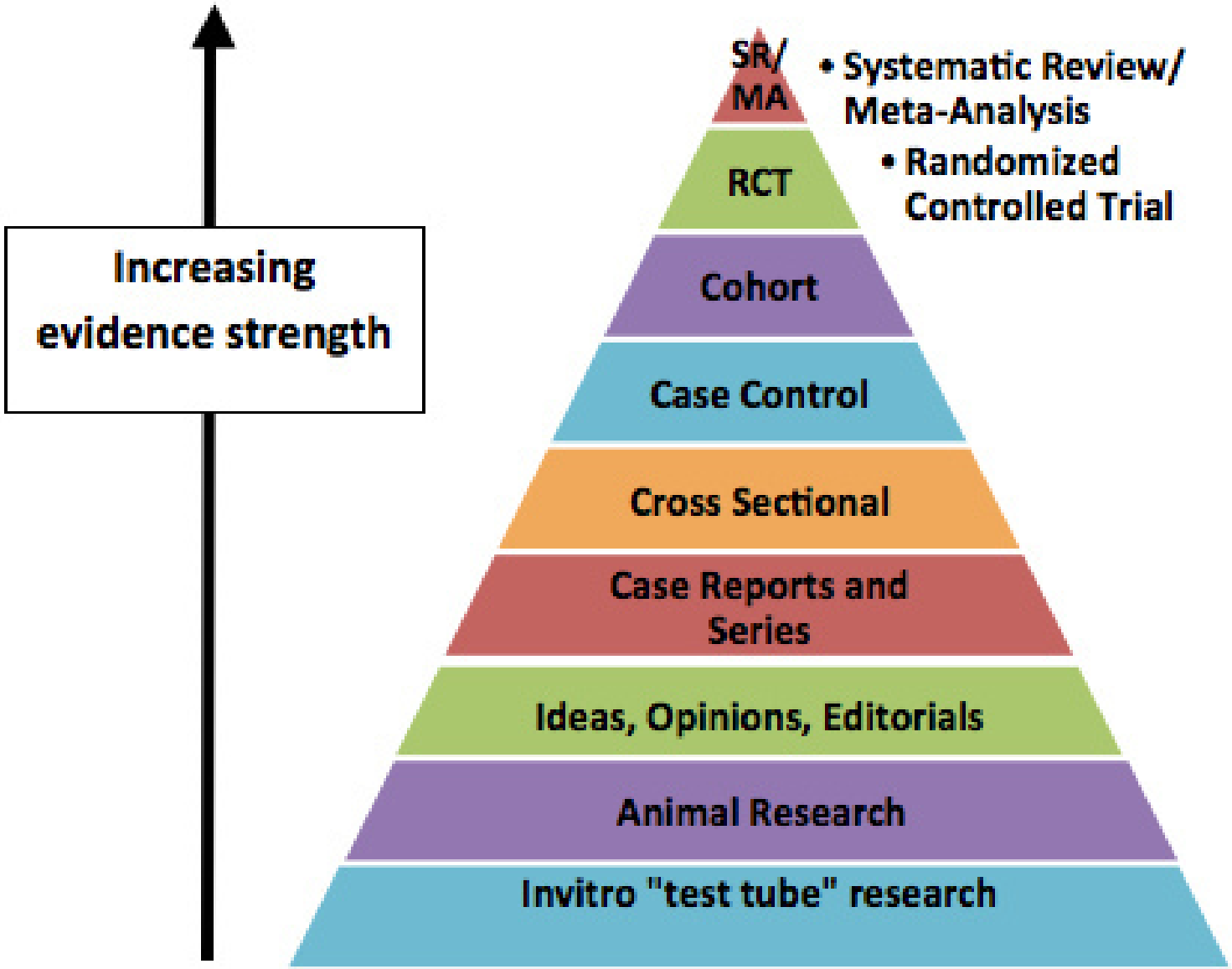


Cross-over design



Factorial design

Strength of evidence



PAUSE

Statistical analyses : independent observations

Choice of statistical test for independent observations

		Outcome variable					
		Nominal	Categorical (>2 Categories)	Ordinal	Quantitative Discrete	Quantitative Non-Normal	Quantitative Normal
Input Variable	Nominal	χ^2 or Fisher's	χ^2	χ^2 trend or Mann-Whitney	Mann-Whitney	Mann-Whitney or log-rank (a)	Student's t test
	Categorical (>2 categories)	χ^2	χ^2	Kruskal-Wallis (b)	Kruskal-Wallis (b)	Kruskal-Wallis (b)	Analysis of variance (c)
	Ordinal (Ordered categories)	χ^2 -trend or Mann-Whitney	(e)	Spearman rank	Spearman rank	Spearman rank	Spearman rank or linear regression (d)
	Quantitative Discrete	Logistic regression	(e)	(e)	Spearman rank	Spearman rank	Spearman rank or linear regression (d)
	Quantitative non-Normal	Logistic regression	(e)	(e)	(e)	Plot data and Pearson or Spearman rank	Plot data and Pearson or Spearman rank and linear regression
	Quantitative Normal	Logistic regression	(e)	(e)	(e)	Linear regression (d)	Pearson and linear regression

Statistical analyses : footnotes

(a) If data are censored.

(b) The Kruskal-Wallis test is used for comparing ordinal or non-Normal variables for more than two groups, and is a generalisation of the Mann-Whitney U test. The technique is beyond the scope of this book, but is described in more advanced books and is available in common software (Epi-Info, Minitab, SPSS).

(c) Analysis of variance is a general technique, and one version (one way analysis of variance) is used to compare Normally distributed variables for more than two groups, and is the parametric equivalent of the Kruskal-Wallis test.

(d) If the outcome variable is the dependent variable, then provided the residuals (see) are plausibly Normal, then the distribution of the independent variable is not important.

(e) There are a number of more advanced techniques, such as Poisson regression, for dealing with these situations. However, they require certain assumptions and it is often easier to either dichotomise the outcome variable or treat it as continuous.

Statistical analyses : dependent observations

Choice of statistical test from paired or matched observation	
Variable	**Test**
Nominal	MeNemar's Test
Ordinal (Ordered categories)	Wilcoxon
Quantitative (Discrete or Non-Normal)	Wilcoxon
Quantitative (Normal*)	Paired ttest
* It is the difference between the paired observations that should be plausibly Normal.	

Matched data : 2 × 2 table

- Like a test/re-test situation, each individual is measured twice
- Also applies to *different* individuals who are not independent : *matched* individuals, siblings, *etc.*

	Test 2 +	Test 2 -	Row total
Test 1 +	a	b	a + b
Test 1 -	c	d	c + d
	a + c	b + d	n

- The null hypothesis of *marginal homogeneity* states that the two marginal probabilities for each outcome are the same : $p_a + p_b = p_a + p_c$ and $p_c + p_d = p_b + p_d$

McNemar's test

- The corresponding null and alternative hypotheses are :

$$H : p_b = p_c$$
$$H : p_b \neq p_c$$

- the *McNemar test statistic* is given by :

$$X^2 = \frac{(b-c)^2}{b+c}$$

- Under the null hypothesis, with a sufficiently large number of discordants (b+c ≥ 25, neither too small), $X^2 \sim \chi^2_1$
- Small sample (exact) analysis : binomial
- Practical application : transmission disequilibrium test for (TDT) testing linkage in the presence of family association

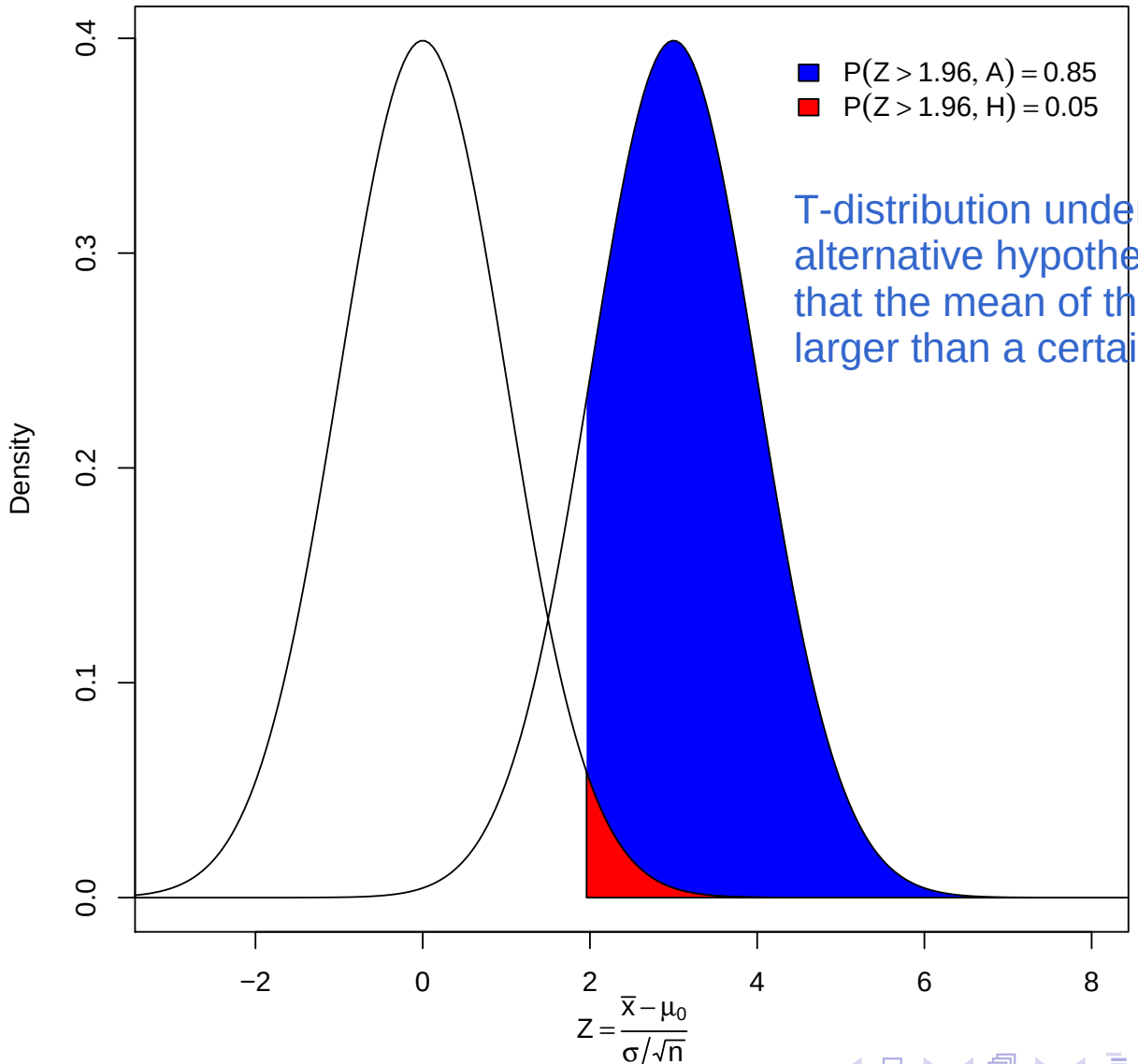
Hypothesis testing

Truth \ Decision	not rejected	rejected
true H	☺ specificity	X Type I error (False +) α
false H	X Type II error (False -) β	☺ Power 1 - β; sensitivity

Power

- We not only want a low false positive rate (α), but also a high *true positive* rate, *i.e.* a high *power* : the power of finding a real effect
- Statistical tests cannot detect a true difference if the sample size is too small compared to the effect size of interest
- In order to calculate/estimate the power for a study, we must specify :
  - the test size (α)
  - the sample size n
  - the effect size d, and
  - the variance σ^2 (or at least an estimate)
- Analogously, we may be interested in finding the sample size n necessary to achieve a given power level

Power : graphically



T-distribution under a specific alternative hypothesis (in this case, that the mean of the population is larger than a certain value).

H:\mu=100 vs A:\mu=108 <- specific alt. hyp. Z ~ N(0,1) X^{crit}= Z*\sigma(n/\sqrt(n))
Z=(x-100)/(\sigma\sqrt(n)) => \mu+Z\cdot\sigma/\sqrt(n) => X^{crit}= 100+1.645(16/\sqrt(16))=106.58
Power=P(X>=106.58|\mu=108)
=P((x-108)/(16\sqrt(16))>=(106.58-108)/(16\sqrt(16))
P(Z>=.3)~=

Power curve : example

- Let X denote the IQ of a randomly selected adult. Also assume that X is normally distributed with unknown mean μ and (known) standard deviation 16.
- We take a random sample of $n = 16$ students, and test the hypotheses :

$$H : \mu = 100$$
$$A : \mu > 100$$

- What is the power of the hypothesis test if the true population mean were $\mu = 108$ (assume $\alpha = 0.05$) ?
- $$\left[n \geq \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\frac{\mu_1 - \mu_2}{\sigma})^2} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\frac{\delta}{\sigma})^2} \right]$$
- For $\alpha = 0.05$ and 80% power ($\beta = 0.20$), then :

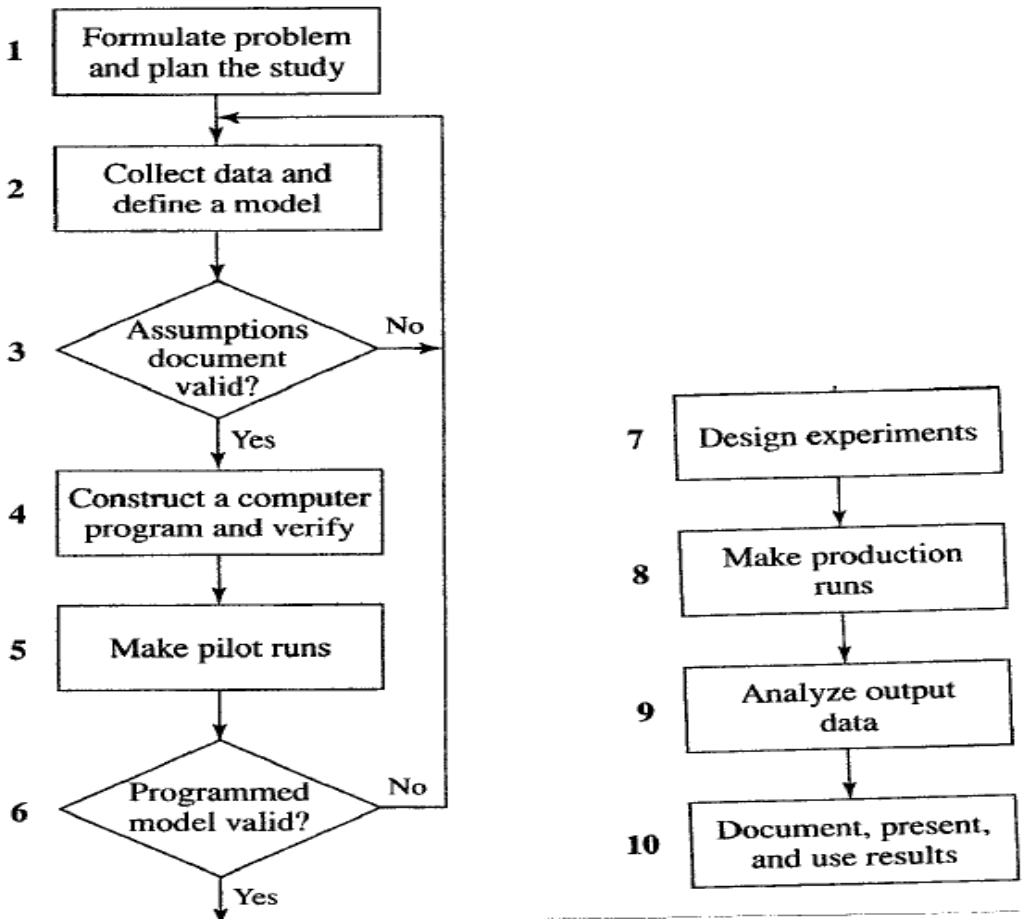
$$\Longrightarrow n = \frac{16}{\Delta^2}, \quad \Delta = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\delta}{\sigma}$$

Basic simulation modeling

- Many/most real-world systems are too complex to study analytically
- Simulation uses computers to *imitate* (*simulate*) real-world processes and study them numerically
- Process of interest is called a *system*
- To study the system, need to make *assumptions* about how it works
- The assumptions form a *model* that is used to try to understand system behavior
- If the system (and corresponding model) are simple enough, could find exact/analytic solution

but only is it's simple enough!

Steps in a simulation study



Simulation advantages

- Might be the *only type* of investigation possible
- Can estimate *performance*
- Can *compare* alternative models
- *Better control* over experimental conditions than possible with actual system experiment
- Can study systems with long time frames, or study system in greater detail

Simulation disadvantages

- Simulation is *stochastic* (i.e. there's a random component in it)
  - need *multiple, independent runs* to produce good estimates
  - choice of input probability distributions
  - random number generator: *set and save the seed* so that simulated values are *reproducible*
- Can be time-consuming
- If the model is wrong, results provide *little useful information* about the actual system

Simulation pitfalls

- Failure to have well-defined *objectives*
- Inappropriate level of *model detail*
- *Misunderstanding* of the simulation by other team members
- Treating the simulation study as a *simple exercise* in computer programming
- Failure to collect good system data

Techniques for increasing model validity and credibility

- Collect *high-quality information* and data on the system
- *Interact* with subject matter experts, managers on a regular basis
- Maintain a document on assumptions *in writing*
- Validate model components *quantitatively*
- Validate *output* from the overall simulation model

Comparing simulated output to real-world observations

- Basic inspection – compare summary parameters
  - Problem: essentially have a ‘sample’ of size 1
- *Correlated* inspection
- *Confidence interval* based on independent data
- Time-series approaches (e.g. spectral analysis)
- *Bootstrapping* hypothesis testing

Power analysis by simulation

- The power calculation formula takes *assumptions* and returns an *analytic solution*
- Since we have computers, it is not necessary to rely on analytic solutions for power analysis
- Program the computer to run the experiment thousands of times then count how frequently the experiment comes up significant
- For any simulation to be *reproducible*, you need to set a *seed* (a place in a very long sequence of random numbers)
- in `R`, the command is `set.seed()`
- As an argument, you give a number, for example 81014 (or whatever your favorite number is!!)
- If interested, you can do some of this during the lab

Power simulation : example R code

```
possible.ns <- seq(from=100, to=2000, by=50) # The sample sizes we'll be considering
powers <- rep(NA, length(possible.ns)) # Empty object to collect simulation estimates
alpha <- 0.05 # Standard significance level
sims <- 500 # Number of simulations to conduct for each N

#### Outer loop to vary the number of subjects ####
for (j in 1:length(possible.ns)){ N <- possible.ns[j] # Pick the jth value for N

  Y0 <- rnorm(n=N, mean=60, sd=20) # control potential outcome
  tau <- 5 # Hypothesize treatment effect
  Y1 <- Y0 + tau # treatment potential outcome
  significant.experiments <- rep(NA, sims) # Empty object to count significant experiments

  #### Inner loop to conduct experiments "sims" times over for each N ####
  for (i in 1:sims){
        Z.sim <- rbinom(n=N, size=1, prob=.5) # Do a random assignment
        Y.sim <- Y1*Z.sim + Y0*(1-Z.sim) # Reveal outcomes according to assignment
        fit.sim <- lm(Y.sim ~ Z.sim) # Do analysis (Simple regression)
        p.value <- summary(fit.sim)$coefficients[2,4] # Extract p-values
        significant.experiments[i] <- (p.value <= alpha) # Determine significance according to
p <= 0.05
        }
  powers[j] <- mean(significant.experiments) # store average success rate (power) for each N
  }
powers
```