Universitat de Barcelona
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

3. Cause
3.3.- Effects

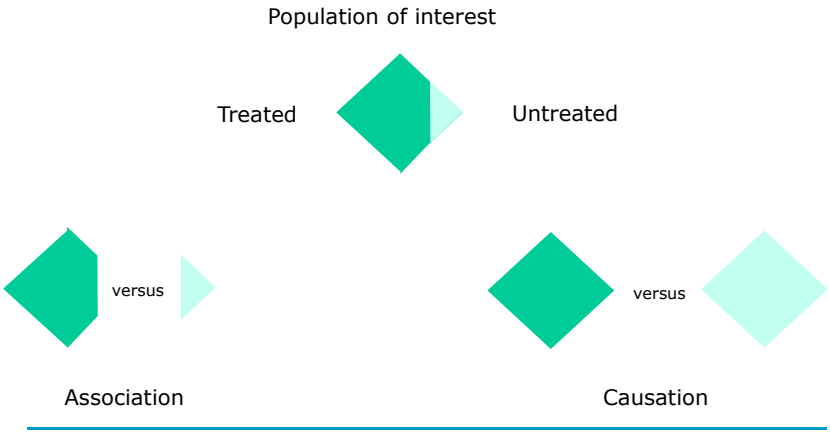
Medical Statistics

José Antonio González y Erik Cobo

Abril 2015

1.- Causal effect

Association versus Causation



a.- Definition

Let: - Two causes X: **t**: alternative **t**reatment in study

c: **c**ontrol

- A population of units $\mathbf{U_i}$
- An outcome **Y** with two possible manifestations:

$\mathbf{Y_i(c)}$: outcome Y in the unit u_i when u_i is allocated to **c**

$\mathbf{Y_i(t)}$: outcome Y in the unit u_i when u_i is allocated to **t**

Example:

DBP of patient 7 if he receives the control **c**: $\mathbf{Y_7(c)}$

DBP of patient 7 if he receives the treatment **t**: $\mathbf{Y_7(t)}$

Causal Effect in a unit i

The effect of the cause 't'
relative to the control 'c'
on the outcome Y
in the individual u_i
is defined as: **t causes the effect $y_i(t) - y_i(c)$**

In the example:
The effect of the cause 'new treatment'
relative to the control
on the outcome DBP
in patient 7
is defined as: **the new treatment causes the effect $y_7(t) - y_7(c)$**

b. Implications of the definition

(i) We define the **effect of a cause** and not the reverse.

Just answer the question prospectively:

¿what is the effect of this cause?

Avoid "competence" with causes of the same effect.

Example: Toxin versus germs.

b. Implications of the definition

(ii) Allocation

We're talking about allocation:

" Outcome when 't' was assigned"

This definition only considers manipulables aspects

It doesn't apply to attributes Z such as gender

Remember:

i) Attributes can help to predict or to characterize the response

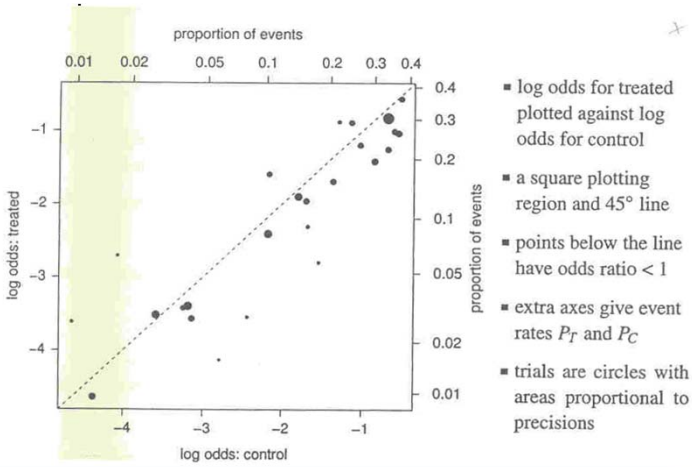
ii) The gender isn't assignable, but a photo and a name (masculine o feminine) can be assigned to a curriculum

iii) Smoking isn't assignable, but the "advice not to smoke" is

iv) Age isn't assignable, but a molecule that stops the aging is

(iii) Relative to other cause: - removes the potential response.

- emulates the 'practice' decision



- log odds for treated plotted against log odds for control
- a square plotting region and 45° line
- points below the line have odds ratio < 1
- extra axes give event rates P_T and P_C
- trials are circles with areas proportional to precisions

b. Implications of the definition

(iv) Potential response

Example: Y is the pain in a scale from 0 to 20

Unit	Potential response		Causal effect
	Y(t)	Y(c)	Y(t)-Y(c)
You	5	15	-10

Unit	Baseline Z	Potential response		Causal effect
		Y(t)-Z	Y(c)-Z	[Y(t)-Z]-[Y(c)-Z]
You	18	-13	-3	-10

Med. Stat. Causal effect

Potential response in a population of 8 cases

Unit	Potential response Y(t)	Y(c)	Causal effect Y(t)-Y(c)
1	14	13	1
2	0	6	-6
3	1	4	-3
4	2	5	-3
5	3	6	-3
6	1	6	-5
7	10	8	2
8	9	8	1
Mean			-2

Changing 'c' by 't'

9

Med. Stat. Causal effect

Only 1 potential response will be observed

Unit	Alloca-tion	Potential response Y(t)	Y(c)	Causal effect Y(t)-Y(c)
1	1	14	~~13~~	?
2	1	0	~~6~~	?
3	1	1	~~4~~	?
4	1	2	~~5~~	?
5	0	~~3~~	6	?
6	0	~~2~~	6	?
7	0	~~10~~	8	?
8	0	~~9~~	8	?
Mean		4.25	7	?

If patients choose their treatment, no rationale to compare 4.25 to 7!!

10

A non observed potential response implies

the fundamental problem of causal inference:

$Y_i(t)$, $Y_i(c)$ can't be observed at once and in the same conditions:

Y(t) can be observed in some units

Y(c) can be observed in others.

Or: $Y_i(t)$ can be observed in a specific conditions

$Y_i(c)$ can be observed in others

Thus, one is observed, but not its complementary.

In summary,

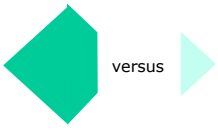
The causal effect in the unit can't be observed

2.- Causal effect on the population

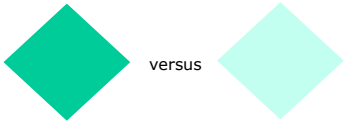
Taking into account the entire population removes the problem because Y(t) can be observed in some units and Y(c) in others.

t causes the (average) effect $E[Y_i(t)] - E[Y_i(c)]$

Association



Causation



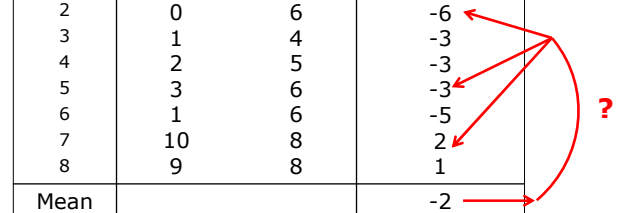
$E(Y_i|t) - E(Y_i|c) \neq E[Y_i(t)] - E[Y_i(c)]$

Med. Stat. Causal effect

If the effect isn't constant over the units,

what cases apply the mean to?

Unit	Potential response Y(t)	Y(c)	Causal effect Y(t)-Y(c)
1	14	13	1
2	0	6	-6
3	1	4	-3
4	2	5	-3
5	3	6	-3
6	1	6	-5
7	10	8	2
8	9	8	1
Mean			-2



Med. Stat. Causal effect

Without constant effect, clinician should allocate best treatment to any case

Unit	Alloca-tion	Potential response Y(t)	Y(c)	Causal effect Y(best) - Y(worst)
1	1	~~14~~	13	-1
2	0	0	~~6~~	-6
3	0	1	~~4~~	-3
4	0	2	~~5~~	-3
5	0	3	~~6~~	-3
6	0	1	~~6~~	-5
7	1	~~10~~	8	-2
8	1	~~9~~	8	-1
	Mean	1.4	9.67	-8.23 ≠ -3

Changing 'best' by 'worst' at any unit

But, how to know…?

Med. Stat. Causal effect

Constant effect (new data)

An assumption for the effect is necessary.

Simplest assumption: **constant and additive effect t**

All units share the same effect

Unit	Potential response Y(t)	Potential response Y(c)	Causal effect Y(t)-Y(c)
1	11	13	-2
2	4	6	-2
3	2	4	-2
4	3	5	-2
5	4	6	-2
6	4	6	-2
7	6	8	-2
8	7	9	-2
Mean			-2

(It's exact because there is no within-case variability in this example.)

Med. Stat. Causal effect

Constant effect: random allocation

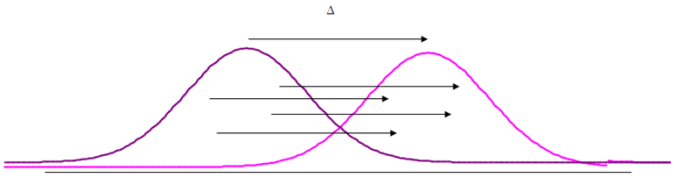
Unit	Alloca-tion	Potential response Y(t)	Potential response Y(c)	Causal effect Y(t) - Y(c)
1	1	11	?	?
2	0	?	6	?
3	0	?	4	?
4	0	?	5	?
5	0	?	6	?
6	0	?	6	?
7	1	6	?	?
8	1	7	?	?
Mean		8	5.4	2.6

Different allocations involve different estimations.

Random allocation: no bias and known variance estimation.

Med. Stat. Causal effect

Constant effect advantages



1) The average effect is pertinent for each unit

[the same dosage can be use for all patients]

2) Implies 'parametric' assumptions: homoscedasticity, same distribution,…

[If not, how to know if there is an unit-treatment interaction?]

3) With random, it solves fundamental problem of causal inferences

Med. Stat. Causal effect

Statistics needs n>1: Group interactions are usually feasible.

Search for interactions

¿what conditions "Z" allow a uniform decision?

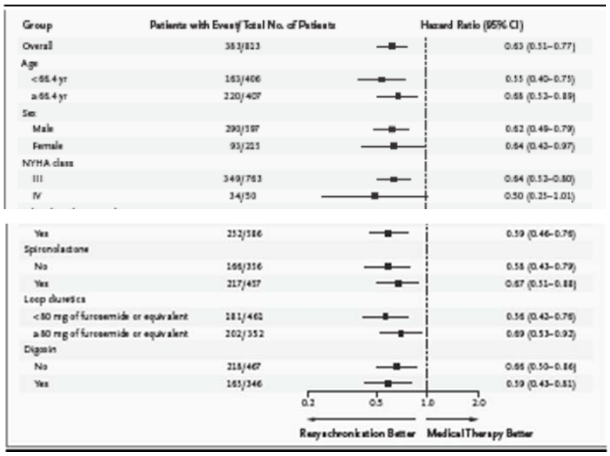
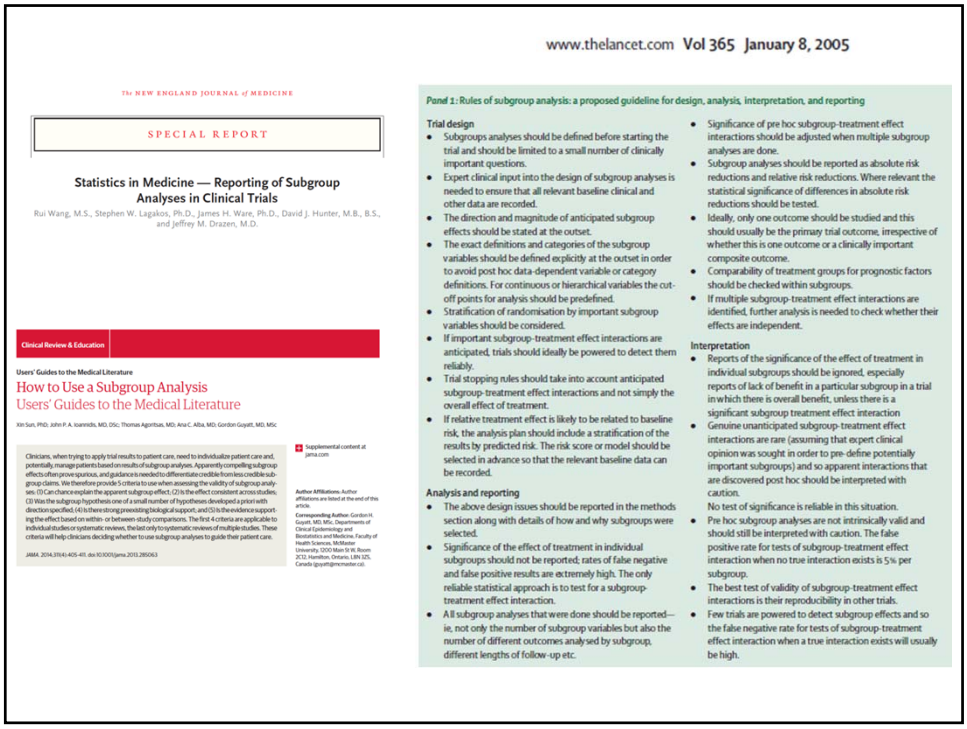


Figure 2 (facing page). Effect of Cardiac Resynchronization on the Primary End Point in Predefined Subgroups.
Hazard ratios and 95 percent confidence intervals (CIs) are shown. The subgroups of age, systolic blood pressure, mitral-regurgitation area (as defined in Table 1), interventricular mechanical delay, ejection fraction, end-systolic volume index, and glomerular filtration rate are divided according to the median value in the study population. All analyses were stratified according to the NYHA class, except the subgroup analysis of NYHA class. To convert values for N-terminal brain natriuretic peptide (NT-BNP) to picomoles per liter, divide by 8.457. For some data (QRS width, for instance), many patients had results at the median value, and this led to some inequality in the sizes of the subgroups. Because of missing baseline data, not all subgroup numbers total 813.

But unit interactions need repeated measures: feasible? Ethical?



www.thelancet.com Vol 365 January 8, 2005

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials

Rui Wang, M.S., Stephen W. Lagakos, Ph.D., James H. Ware, Ph.D., David J. Hunter, M.B., B.S., and Jeffrey M. Drazen, M.D.

Clinical Review & Education

Users' Guides to the Medical Literature

How to Use a Subgroup Analysis
Users' Guides to the Medical Literature

Xin Sun, PhD; John P. A. Ioannidis, MD, DSc; Thomas Agoritsas, MD; Ana C. Alba, MD; Gordon Guyatt, MD, MSc

Clinicians, when trying to apply trial results to patient care, need to individualize patient care and, potentially, manage patients based on results of subgroup analyses. Apparently compelling subgroup effects often prove spurious, and guidance is needed to differentiate credible from less credible subgroup claims. We therefore provide 5 criteria to use when assessing the validity of subgroup analyses. (1) Can chance explain the apparent subgroup effect; (2) is the effect consistent across studies; (3) Was the subgroup hypothesis one of a small number of hypotheses developed a priori with direction specified; (4) is there strong preexisting biological support; and (5) is the evidence supporting the effect based on within- or between-study comparisons. The first 4 criteria are applicable to individual studies or systematic reviews, the last only to systematic reviews of multiple studies. These criteria will help clinicians deciding whether to use subgroup analyses to guide their patient care.

JAMA. 2014;311(4):405-411. doi:10.1001/jama.2013.285063

Supplemental content at jama.com

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Gordon H. Guyatt, MD, MSc, Departments of Clinical Epidemiology and Biostatistics and Medicine, Faculty of Health Sciences, McMaster University, 1200 Main St W, Room 2C12, Hamilton, Ontario, L8N 3Z5, Canada (guyatt@mcmaster.ca).

Panel 1: Rules of subgroup analysis: a proposed guideline for design, analysis, interpretation, and reporting

Trial design

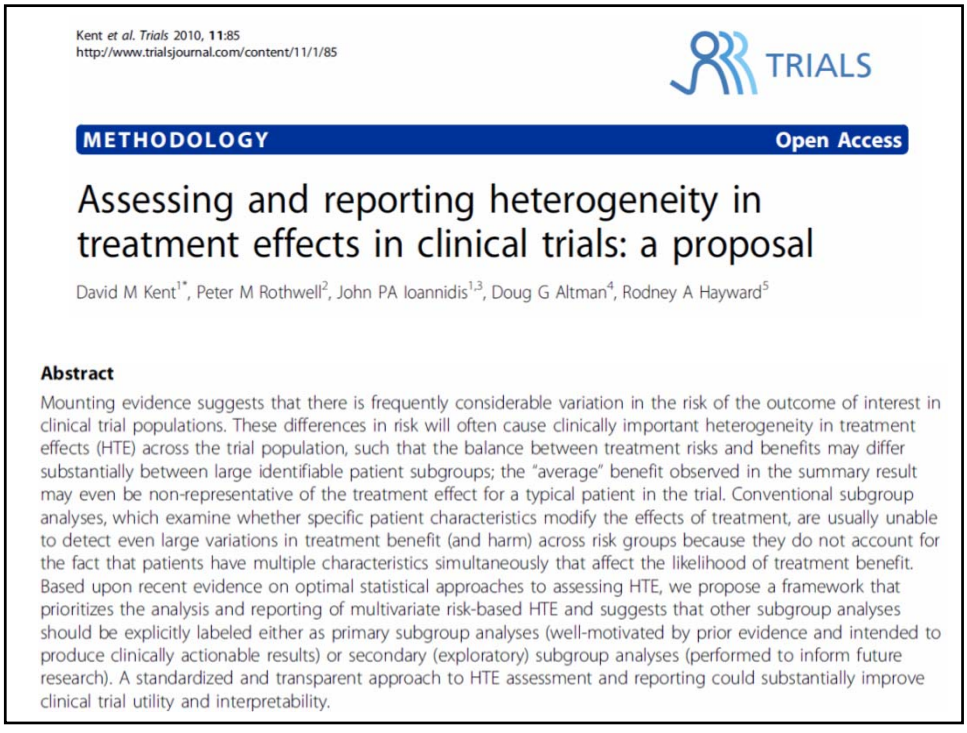
- Subgroups analyses should be defined before starting the trial and should be limited to a small number of clinically important questions.
- Expert clinical input into the design of subgroup analyses is needed to ensure that all relevant baseline clinical and other data are recorded.
- The direction and magnitude of anticipated subgroup effects should be stated at the outset.
- The exact definitions and categories of the subgroup variables should be defined explicitly at the outset in order to avoid post hoc data-dependent variable or category definitions. For continuous or hierarchical variables the cut-off points for analysis should be predefined.
- Stratification of randomisation by important subgroup variables should be considered.
- If important subgroup-treatment effect interactions are anticipated, trials should ideally be powered to detect them reliably.
- Trial stopping rules should take into account anticipated subgroup-treatment effect interactions and not simply the overall effect of treatment.
- If relative treatment effect is likely to be related to baseline risk, the analysis plan should include a stratification of the results by predicted risk. The risk score or model should be selected in advance so that the relevant baseline data can be recorded.

Analysis and reporting

- The above design issues should be reported in the methods section along with details of how and why subgroups were selected.
- Significance of the effect of treatment in individual subgroups should not be reported; rates of false negative and false positive results are extremely high. The only reliable statistical approach is to test for a subgroup-treatment effect interaction.
- All subgroup analyses that were done should be reported—ie, not only the number of subgroup variables but also the number of different outcomes analysed by subgroup, different lengths of follow-up etc.
- Significance of pre hoc subgroup-treatment effect interactions should be adjusted when multiple subgroup analyses are done.
- Subgroup analyses should be reported as absolute risk reductions and relative risk reductions. Where relevant the statistical significance of differences in absolute risk reductions should be tested.
- Ideally, only one outcome should be studied and this should usually be the primary trial outcome, irrespective of whether this is one outcome or a clinically important composite outcome.
- Comparability of treatment groups for prognostic factors should be checked within subgroups.
- If multiple subgroup-treatment effect interactions are identified, further analysis is needed to check whether their effects are independent.

Interpretation

- Reports of the significance of the effect of treatment in individual subgroups should be ignored, especially reports of lack of benefit in a particular subgroup in a trial in which there is overall benefit, unless there is a significant subgroup treatment effect interaction
- Genuine unanticipated subgroup-treatment effect interactions are rare (assuming that expert clinical opinion was sought in order to pre-define potentially important subgroups) and so apparent interactions that are discovered post hoc should be interpreted with caution.
  No test of significance is reliable in this situation.
- Pre hoc subgroup analyses are not intrinsically valid and should still be interpreted with caution. The false positive rate for tests of subgroup-treatment effect interaction when no true interaction exists is 5% per subgroup.
- The best test of validity of subgroup-treatment effect interactions is their reproducibility in other trials.
- Few trials are powered to detect subgroup effects and so the false negative rate for tests of subgroup-treatment effect interaction when a true interaction exists will usually be high.



Kent et al. Trials 2010, 11:85
http://www.trialsjournal.com/content/11/1/85

TRIALS

METHODOLOGY Open Access

Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal

David M Kent[1*], Peter M Rothwell[2], John PA Ioannidis[1,3], Doug G Altman[4], Rodney A Hayward[5]

Abstract

Mounting evidence suggests that there is frequently considerable variation in the risk of the outcome of interest in clinical trial populations. These differences in risk will often cause clinically important heterogeneity in treatment effects (HTE) across the trial population, such that the balance between treatment risks and benefits may differ substantially between large identifiable patient subgroups; the "average" benefit observed in the summary result may even be non-representative of the treatment effect for a typical patient in the trial. Conventional subgroup analyses, which examine whether specific patient characteristics modify the effects of treatment, are usually unable to detect even large variations in treatment benefit (and harm) across risk groups because they do not account for the fact that patients have multiple characteristics simultaneously that affect the likelihood of treatment benefit. Based upon recent evidence on optimal statistical approaches to assessing HTE, we propose a framework that prioritizes the analysis and reporting of multivariate risk-based HTE and suggests that other subgroup analyses should be explicitly labeled either as primary subgroup analyses (well-motivated by prior evidence and intended to produce clinically actionable results) or secondary (exploratory) subgroup analyses (performed to inform future research). A standardized and transparent approach to HTE assessment and reporting could substantially improve clinical trial utility and interpretability.

Table 4 Checklist for Reporting on Subgroup Analyses & Heterogeneity in Treatment Effects

1. Evaluate and report on the distribution of risk in the overall study population and in the separate treatment arms of the study by using a risk prediction model or index.

- Report on the distribution of predicted risk (or risk score) in the study population overall and by treatment arm.
- Risk reporting should allow readers to assess the full distribution of the study population either graphically (e.g., histograms or box & whiskers plots) or by including information on the mean, standard deviation, median and interquantile ranges.

2. Primary subgroup analyses should include reporting how relative and absolute risk reduction varies in a risk-stratified analysis.

- The risk prediction model should be pre-specified (i.e., fully specified before *any* analysis of treatment-effect has begun) and preferably externally developed.
- Both absolute and relative risk reductions must be reported.

3. Any additional primary subgroup analysis should be pre-specified and limited to patient attributes with strong a priori pathophysiological or empirical justification.

- All primary subgroup comparisons must be pre-specified.
- Prespecification should include all aspects of the subgroup analysis, including threshold values for continuous or ordinal variables where these are used.
- All primary subgroup analyses must be justified based upon pathophysiological or empirical evidence that this factor modifies treatment effects.

4. Conduct and report on secondary (exploratory) subgroup analyses separately from primary subgroup comparisons.

- Secondary subgroup analyses must be reported separately from primary subgroup analyses and clearly labeled as exploratory (potential useful for hypothesis generation and informing future research, but having little or no immediate relevance to patient care).

5. All analyses conducted must be reported and statistical testing of HTE should be done using appropriate methods (such as interaction terms) and avoiding overinterpretation.

- Reporting must include results for all subgroup analyses conducted and the paper must state that primary subgroup analyses conducted were pre-specified and reported.
- Statistical comparisons should be limited to reporting for statistical significance of treatment heterogeneity between subgroups using interaction terms. (Testing for the significance of a treatment effect within a subgroup is inappropriate due to poor statistical power).
- Statistical comparisons should be corrected for the number of primary subgroup analyses performed.

Med. Stat. Causal effect

¿Is necessary another assumption about the effect?

You take I take	t t	c t	t c	c c
You=1	$Y_1(t,t)=\mathbf{0}$	$Y_1(c,t)=\mathbf{100}$	$Y_1(t,c)=\mathbf{0}$	$Y_1(c,c)=\mathbf{100}$
I=2	$Y_2(t,t)=\mathbf{0}$	$Y_2(c,t)=\mathbf{50}$	$Y_2(t,c)=\mathbf{75}$	$Y_2(c,c)=\mathbf{100}$

In **your** case, the effect is independent of **my** allocation

If **I take**: $Y_1(t,t) - Y_1(c,t) = \mathbf{0-100} = \mathbf{-100}$

If **I don't take**: $Y_1(t,c) - Y_1(c,c) = \mathbf{0-100} = \mathbf{-100}$

In **my** case, the effect depends on **your** allocation

If **you take**: $Y_2(t,t) - Y_2(t,c) = \mathbf{0 - 75} = \mathbf{-75}$

If **you don't take**: $Y_2(c,t) - Y_2(c,c) = \mathbf{50-100} = \mathbf{-50}$

The **assumption** of the **effect independence to other allocations** is also needed

22

3.- Adjusted causal effect

The effect could be constant under some conditions/attributes Z:

t causes the adjusted (average) effect E[Y(t)|z] - E[Y(c)|z]

Example: increase in the adjusted DBP,
when the treatment C is replaced by the treatment T
in a patient with ***fixed*** age and center.

Conditioning, adjusting by Z…

1) Facilitates the homogeneity of effect assumption
2) Reduces outcome/response variability

→ higher efficiency, accuracy and power (lower estimator SE)

3) Controls possible confounders

[Always: use **previous** and **known** outcomes **without measurement error**.]

Options for the adjustment. Adapted from Kleimbaum et al. Epidemiologic Research, 1982

Option	Stage*	Name	Advantages	Inconvenients
Restriction	Design	Eligibility criteria	Complete control Inexpensive Easy to design Easy to analyze	Reduce generalizability Limited number of variables May have "residual confounding"
	Analysis	One subgroup analysis		
Subgroups analysis	Design	Blocks (matching)	More power or efficiency Without assumptions Direct Easy calculations	Flexibility Cost Case dispersion over stratums Different stratifications More difficult to summarize
	Analysis	Stratification (matching)		
Statistical modeling	Design	Modeling	Feasible with small numbers Provides "smoothing" Provides individual "prediction of risk" Allows continuous variables Allows several exposure variables	A lot of assumptions Model choice difficult Predictors choice difficult Interpretation Software parameterization
	Analysis	Covariance Regression		
Overall	Design	Minimization	Simultaneous for several variables Generalizability is not lost	Logistics are sophisticated
	Analysis	Optimal pair		

*Only an adjustment specified in the protocol allows a confirmatory interpretation.
Adjustment in the analysis phase doesn't allow to correct design flaws.