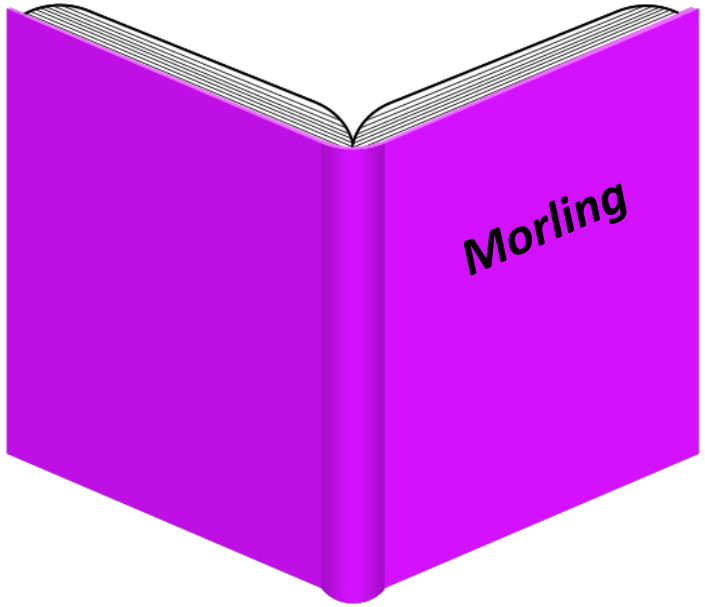




Research Methods and Statistics

Lecture 14: Clinical trials

Riet van Bork

Today

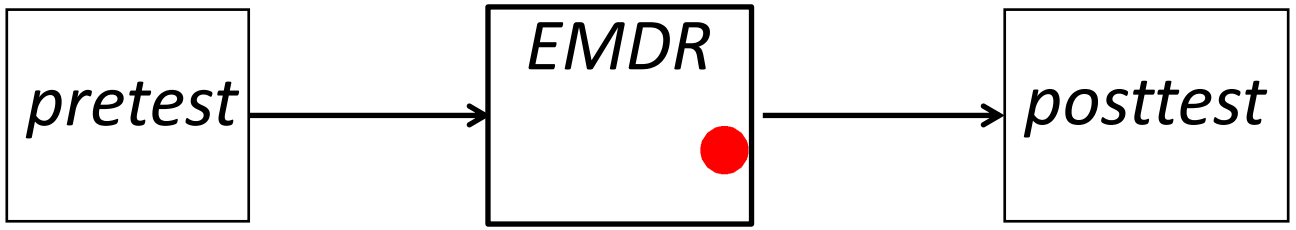
1. **Large *N* clinical trials**
2. Small *N* clinical trials
3. Quasi-experiments

Eye Movement Desensitization and Reprocessing (Shapiro, 1989)



Example today: Does EMDR cause improvement in depression?

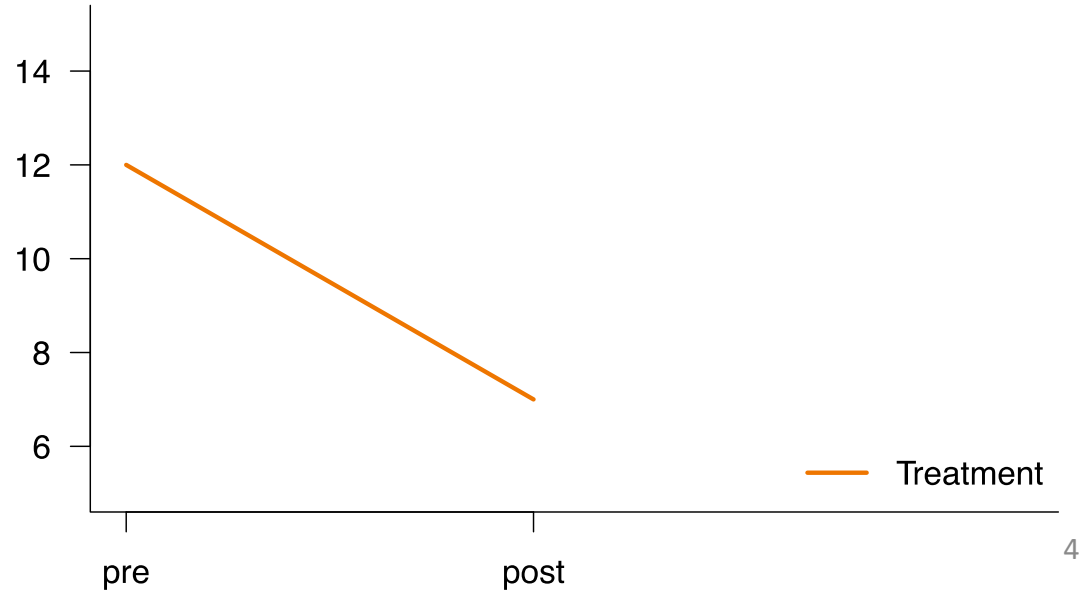
Design #1:
One-group pretest/posttest design



Only consider the point estimates and ignore sampling variability. So, any difference in scores is a difference at the population level.

In that case: there is a decrease in depression score!

But, “EMDR leads to less depression” is a causal claim… so what about internal validity?



Threats to internal validity

Internal validity: *Are there alternative explanations for the statistical effect?*

1. Maturation/spontaneous remission
2. History threat
3. Regression to the mean
4. Non-specific effect (*not in book!*)
7. Placebo
8. Observer bias
9. Demand characteristics
10. Attrition / subject loss
11. Testing threat
12. Instrumentation threat

Maturation threat/spontaneous remission

- Patients improve just because time passes
- *Spontaneous remission:* patients improve without known reason



Source: needpix.com

History threat

- An external factor *independent of the study* that influences most participants

- For example:
  - The weather: If the pretest took place in the winter and the posttest in spring, depressed emotions are reduced because of that
  - Covid: If the pretest took place when there was still a national lockdown, and the posttest was after the lockdown was over, depressed emotions are reduced because of that

Regression (to the mean)

- Participants/Patients with extreme scores are typically less extreme at another measurement
  - From the book:
    - In the 2014 World Cup, Germany beats Brazil 7-1
    - You knew already: next Germany game will not have that many goals



theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

Sport › Football › World Cup 2014

worldcup2014

Disbelief as Germany break hosts Brazil's hearts in 7-1 defeat

Scuffles on Copacabana beach as Brazil faces one of its most traumatic football moments since 1950 World Cup final defeat

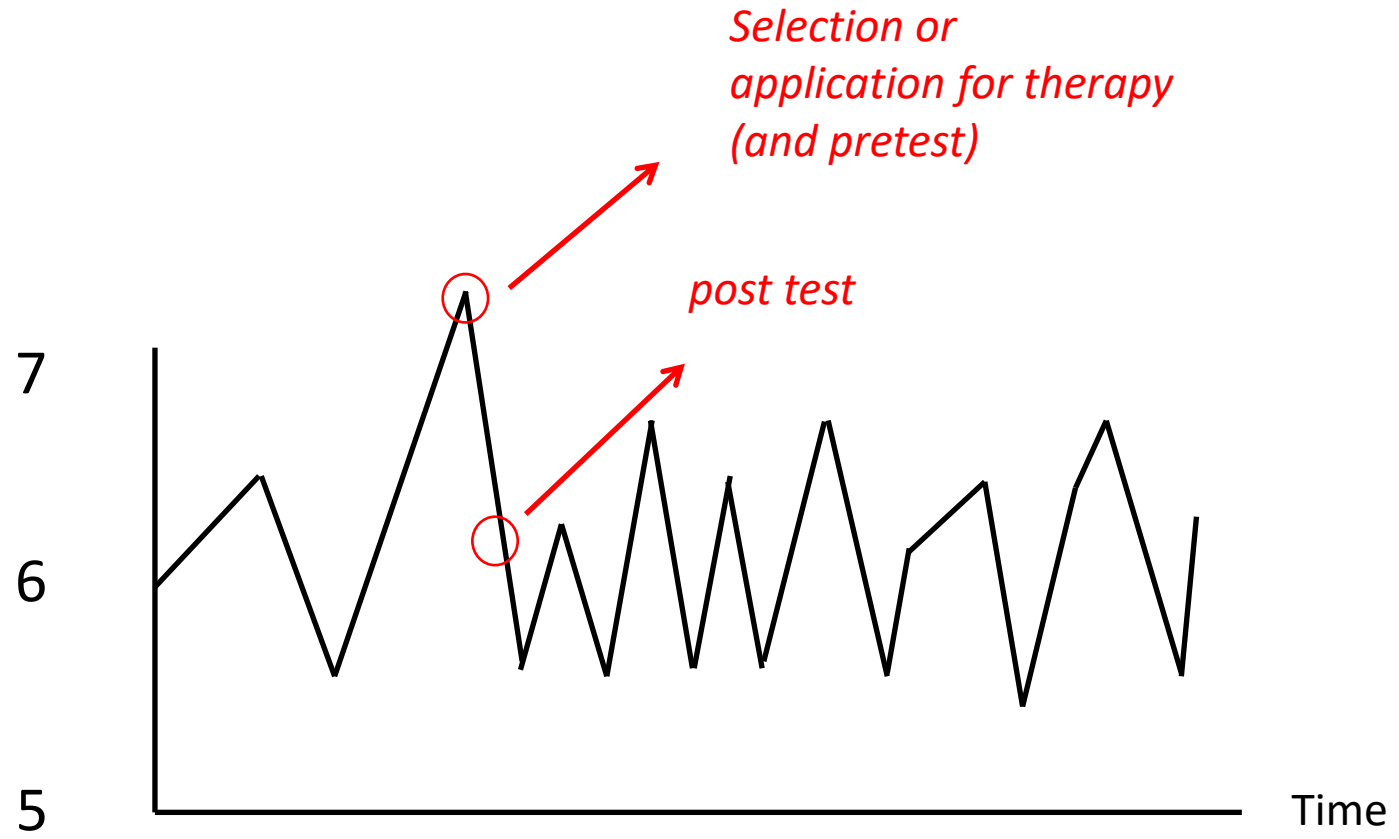
Jonathan Watts in Rio de Janeiro
The Guardian, Wednesday 9 July 2014

Brazil's David Luiz cries after his team's defeat to Germany in the World Cup semi-final. Photograph: Eddie Keogh/Reuters

Photo adapted from: https://www.flickr.com/photos/wolfgangkuhnle/14423600557

Regression (to the mean)

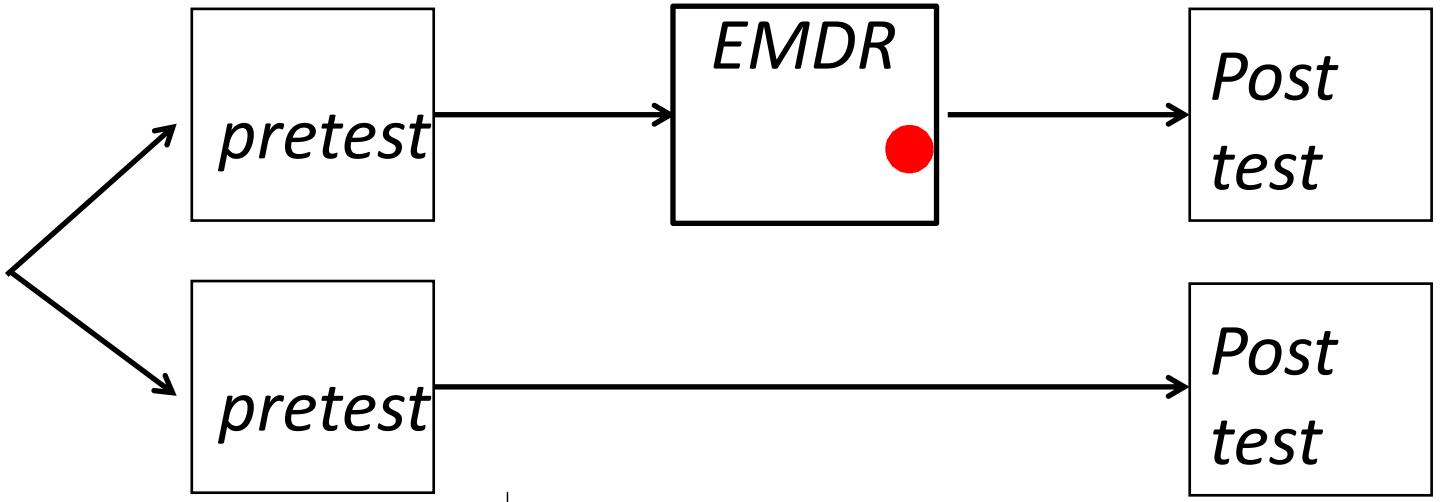
Eg., negative mood fluctuates over time



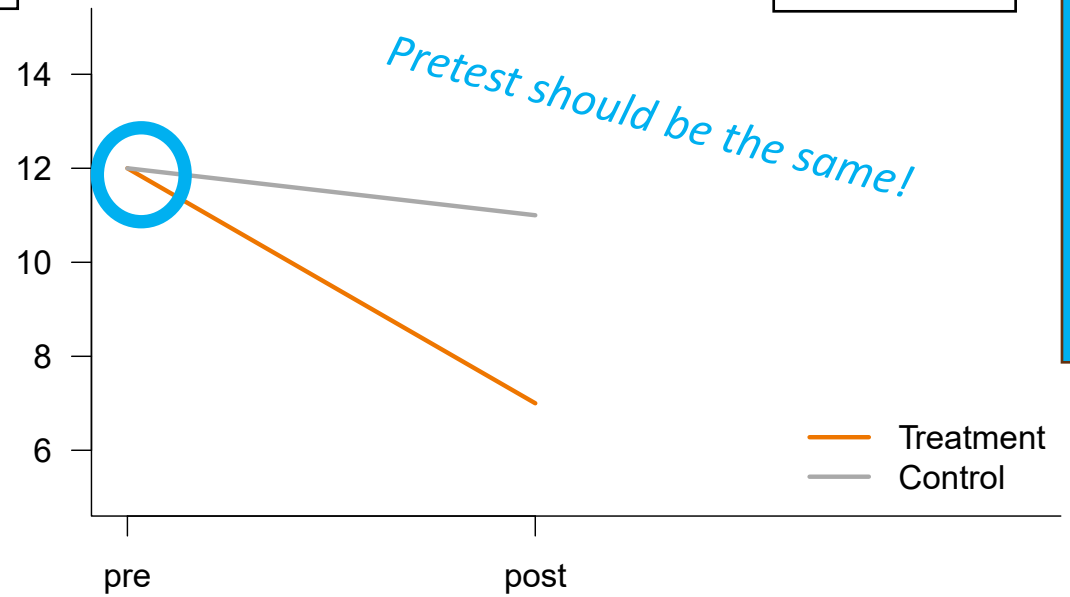
Design #2:
Pretest/posttest design with control group



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
Non-specific effect (*not in book!*)
Placebo
Observer bias
Demand characteristics
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~



If pretest is not the same, you cannot draw conclusions on any of the explanations for the statistical effect

Testing & instrumentation threat

- Testing threats: effects of being tested twice
  - E.g., practice effects, memory effects

    →accounted for by a control group
- Instrumentation
  - The measurement instruments differ between the pre and post test
    - E.g., on post test a more easy arithmatic test is being used
    - E.g., the EEG is calibrated slightly different for the post test

  →accounted for by a control group

  →use the same instrument/calibration or make tests comparable
- Instrumentation vs testing threat
  - Instrumentation threat means the measurement instrument has changed from time 1 to time 2, whereas with testing threat the participants change from time 1 to 2.

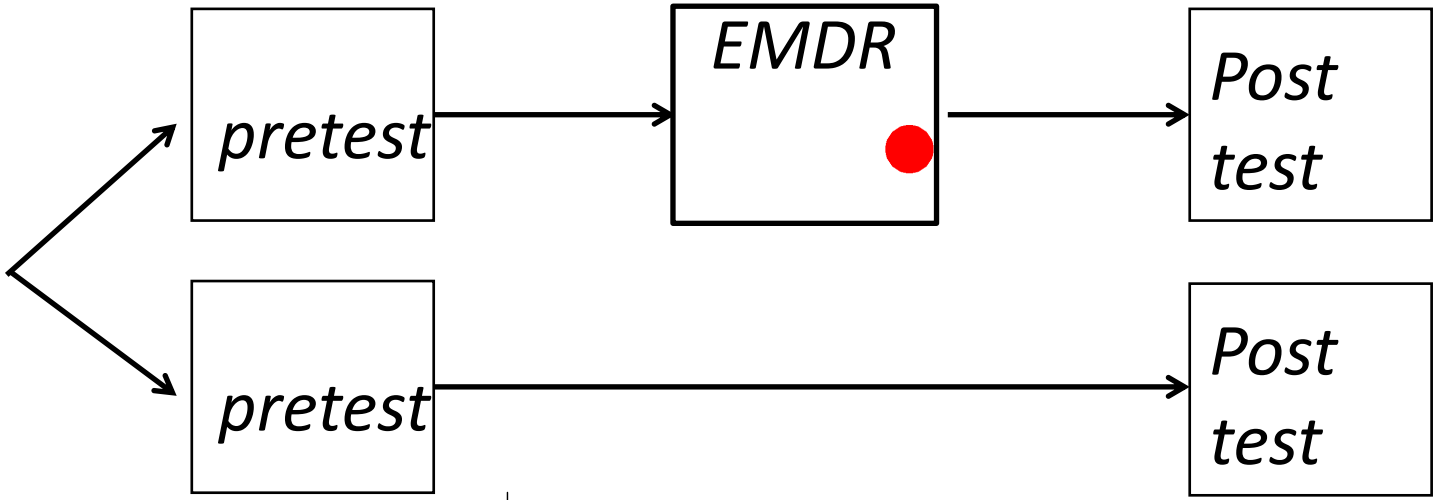


Image source: pixabay.com

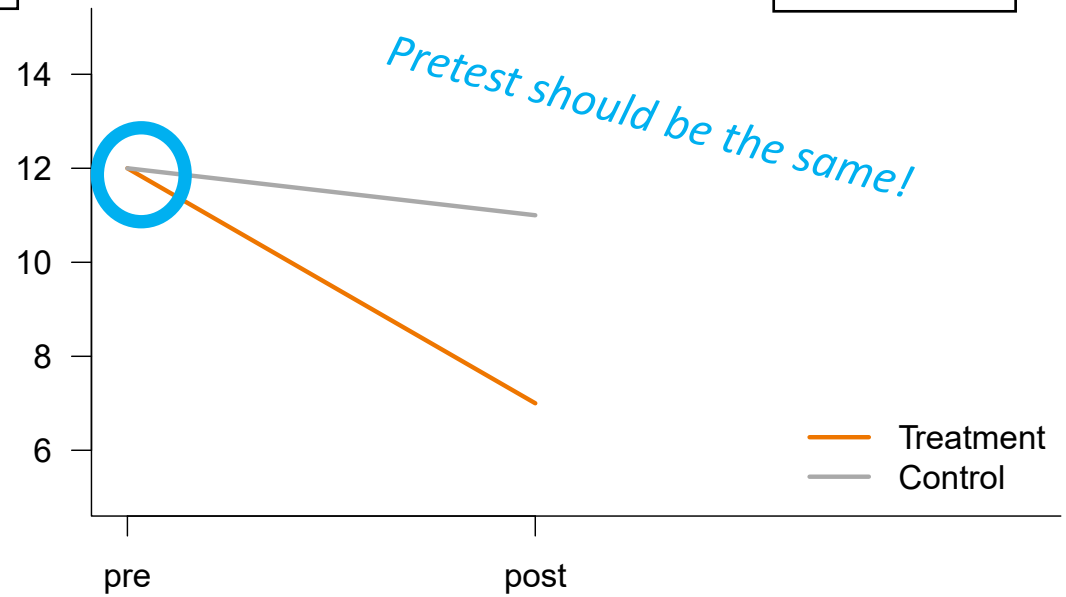
Design #2:
Pretest/posttest design with control group



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
Non-specific effect (*not in book!*)
Placebo
Observer bias
Demand characteristics
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~



Non-specific effects

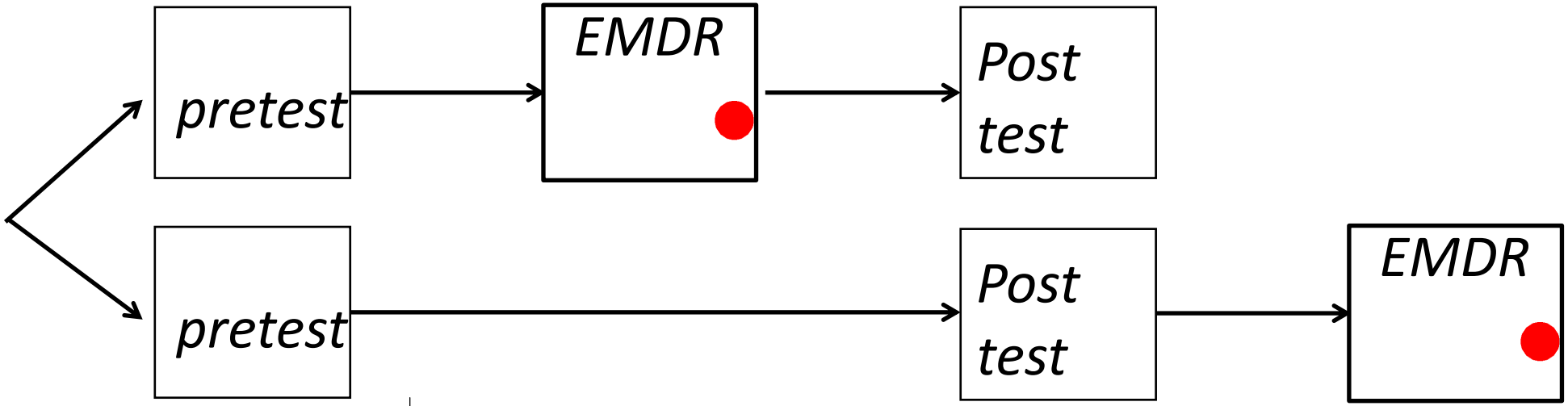
Effects that are not due to the treatment, but due to the expectation of being treated

- Examples:
  - Subjects starts healthy eating habits which improves general wellbeing
  - Subjects seek out social support which improves mood
  - Subjects start physical exercising
  - …

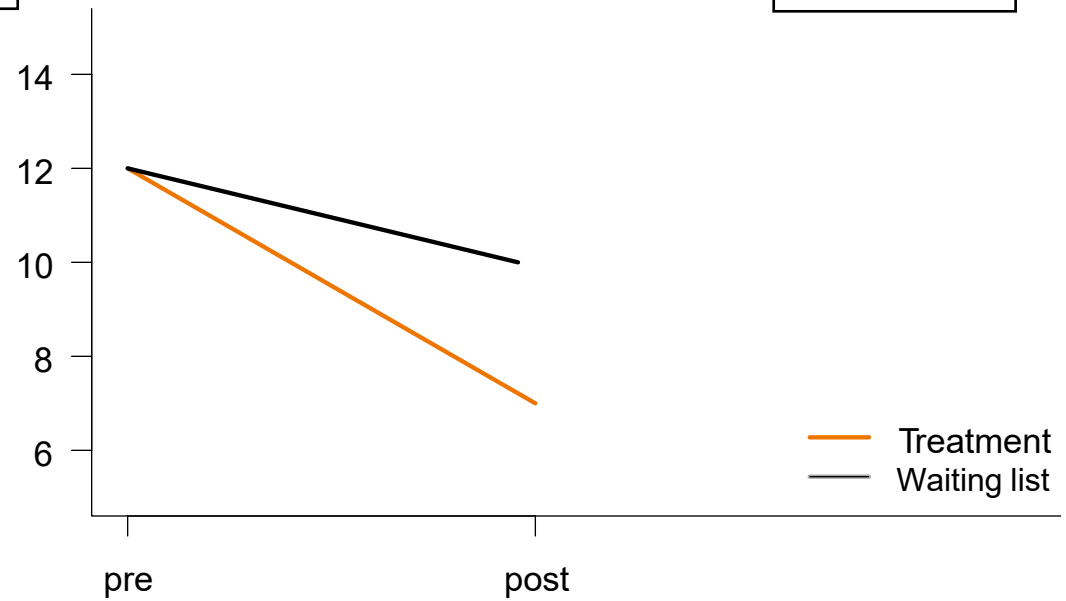
Design #3:
Pretest/posttest design with waiting list control group



Photo from commons.wikimedia.org



- ~~Maturation/spontaneous remission~~
- ~~History threat~~
- ~~Regression to the mean~~
- ~~Non-specific effect (*not in book!*)~~
- Placebo
- Observer bias
- Demand characteristics
- Attrition / subject loss
- ~~Testing threat~~
- ~~Instrumentation threat~~



Placebo effect: The “true” effect of pain killers

- Study done by Amanzio (2001)
- Patients were informed that they would receive pain relieving medicine until their pain was reduced by 50%
- Self-report on a pain scale (“rate your pain”)
- Two groups:
  - “Open”
  - “Hidden”

Open



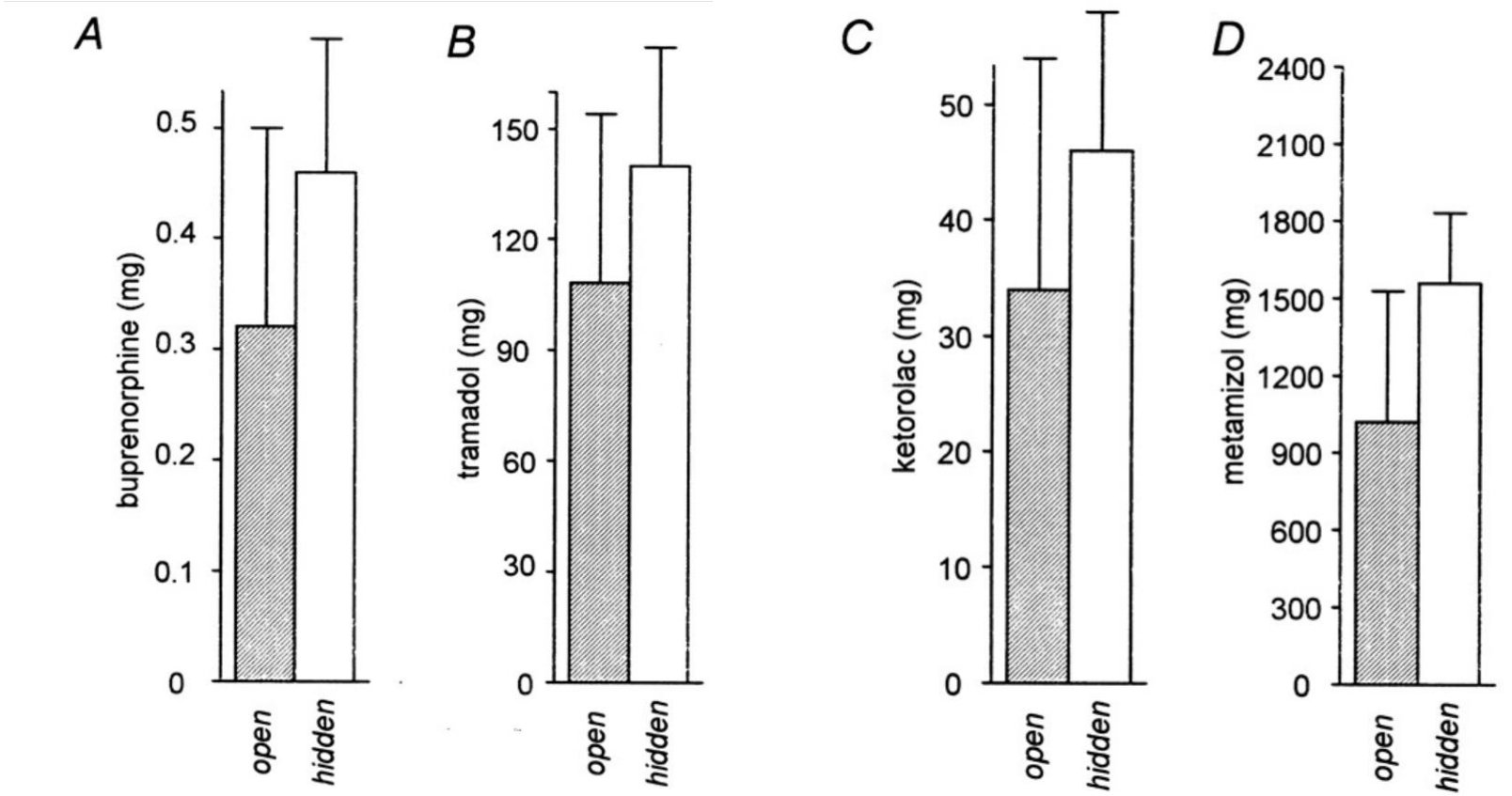
Source: pixabay.com

Hidden



Source: wikipedia.nl

Results



AD_{50} of: (A), buprenorphine; (B), tramadol; (C), ketorolac; and (D), metamizol, obtained by means of either open or hidden infusions in postoperative patients. Note that in the hidden conditions, the AD_{50} increased whereas the SD decreased.

Source: Amanzio, Martina[a]; Pollo, Antonella[a]; Maggi, Giuliano[b]; Benedetti, Fabrizio[a,c,*]. Response variability to analgesics: a role for non-specific activation of endogenous opioids. Pain 90(3):p 205-215, February 2001. | DOI: 10.1016/S0304-3959(00)00486-3

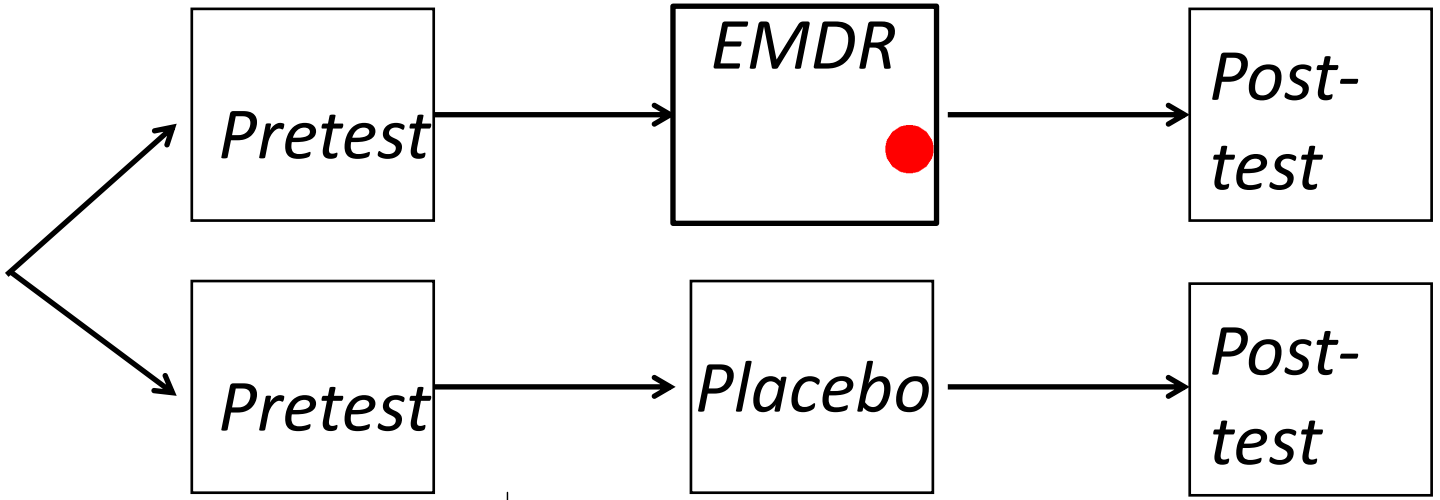
Non-specific effects versus Placebo effect

- Non-specific effects occur when you believe that you ***will be*** treated
- Placebo effects occur when you think you ***are being*** treated
- The other previous effects (maturation, history threat & regression to the mean) occur independent of the participant's/patient's belief

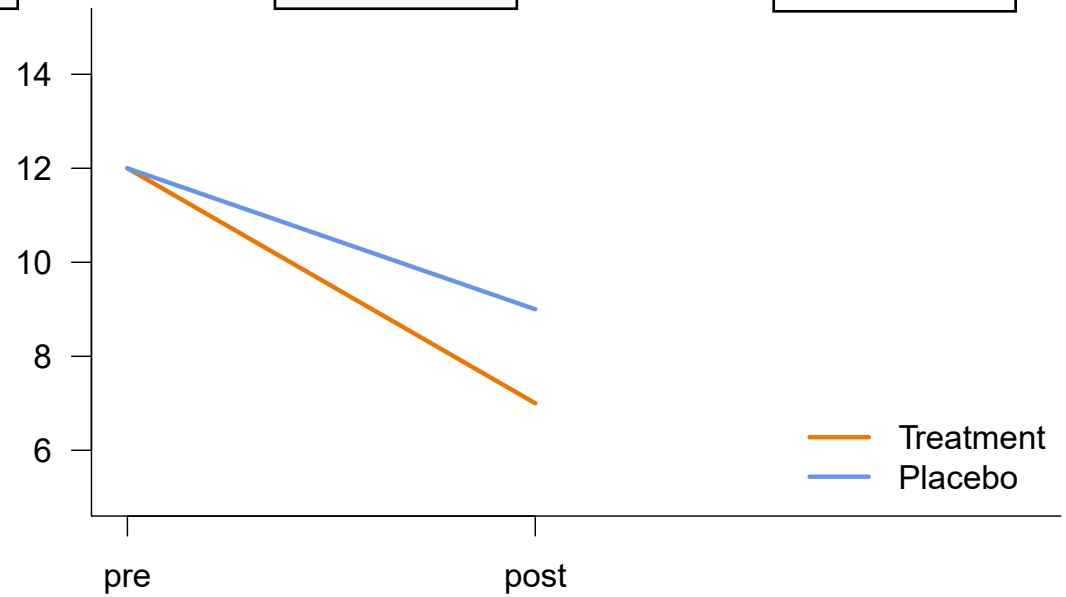
Design #4:
Pretest/posttest design with placebo group



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
~~Non-specific effect (*not in book!*)~~
~~Placebo~~
Observer bias
Demand characteristics
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~



Placebo in clinical psychology

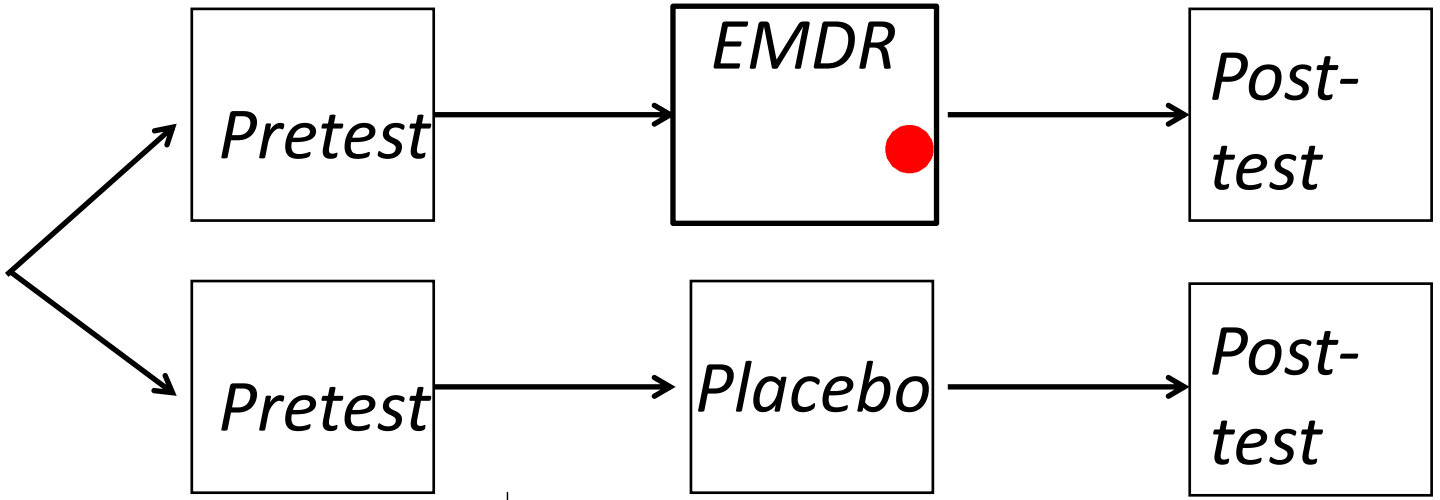
- What is “fake” treatment?
  - Something that does not use clinical psychological techniques
    - E.g., sometimes placebo psychotherapy is simply talking to a friendly listener but in an unstructured conversation (no therapeutic structure)
  - → difficult in practice

- Traditional treatment
  - Instead of comparing to placebo group sometimes it is decided to compare to traditional treatment
  - Research question changes
    - “Is the new therapy better than the traditional therapy?”
  - Traditional therapy is sometimes called ***Positive control group*** (this is different from placebo)

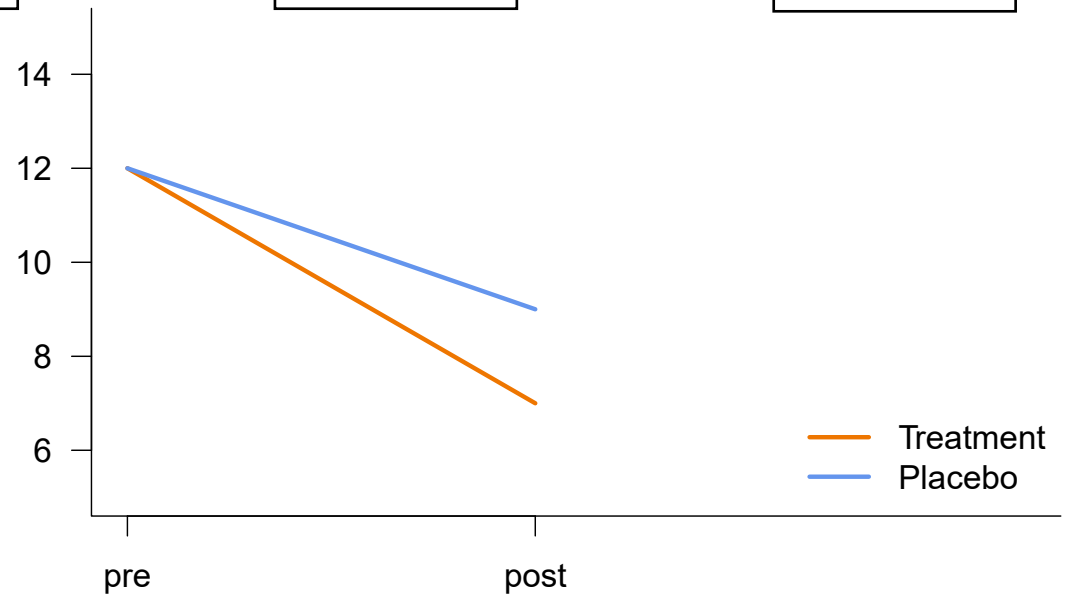
Design #4:
Pretest/posttest design with placebo group



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
~~Non-specific effect (*not in book!*)~~
~~Placebo~~
Observer bias
Demand characteristics
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~



Observer effect & bias and Demand characteristics

- Observer bias (see also Ch6):
  - The **observer** may see differences between the conditions that are not actually there
- Observer effect (see also Ch6):
  - The **observer** may treat participants differently depending on the condition they're in and as such change their behavior to match the researcher's expectations

- Demand characteristics:
  - The **participant** may guess what the study is about and have the tendency to behave according to the research hypothesis

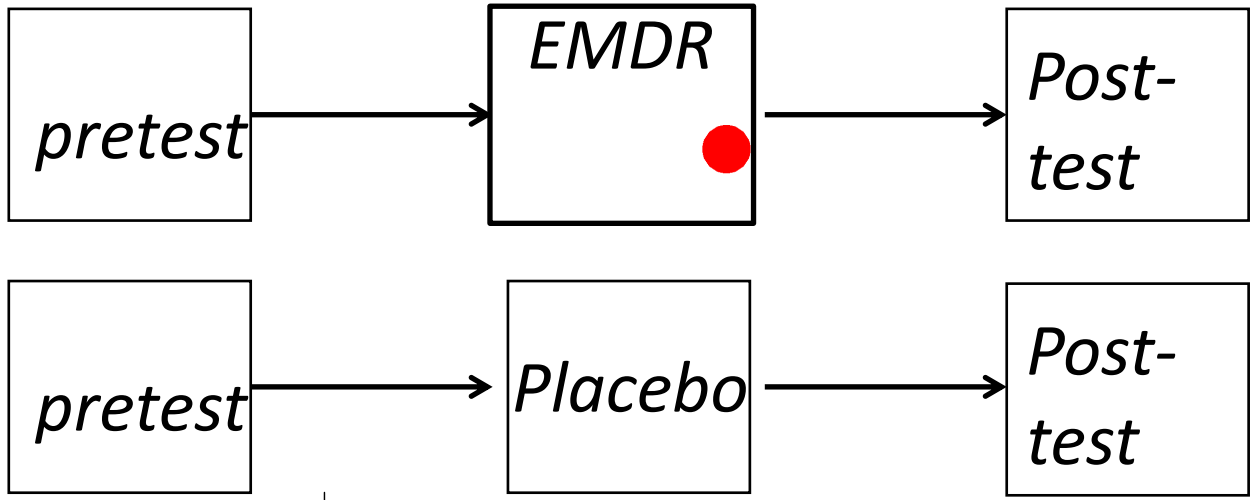
→Double-blind studies (blind to participants and researcher)
or masked design (blind to researcher not to participants)

Design #5:

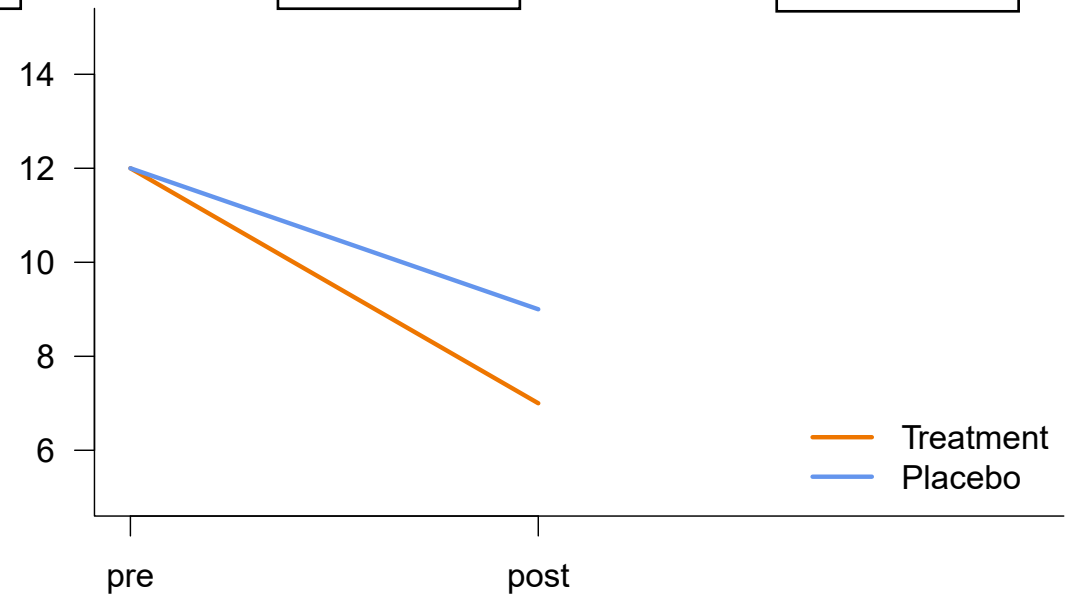
Pretest/posttest design with '*masked design* placebo control'



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
~~Non-specific effect (*not in book!*)~~
~~Placebo~~
~~Observer bias~~
Demand characteristics
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~

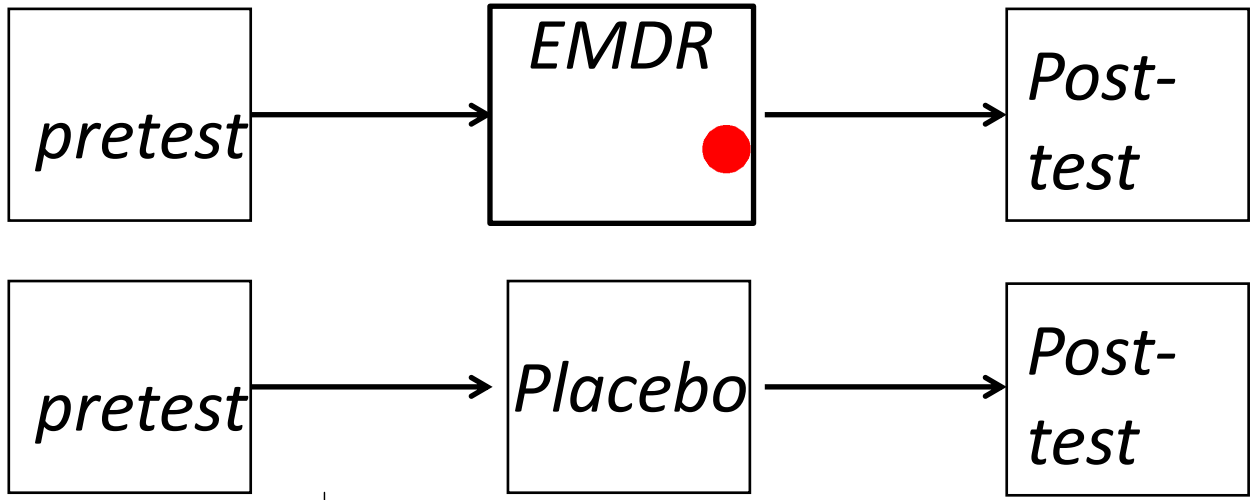


Design #5:

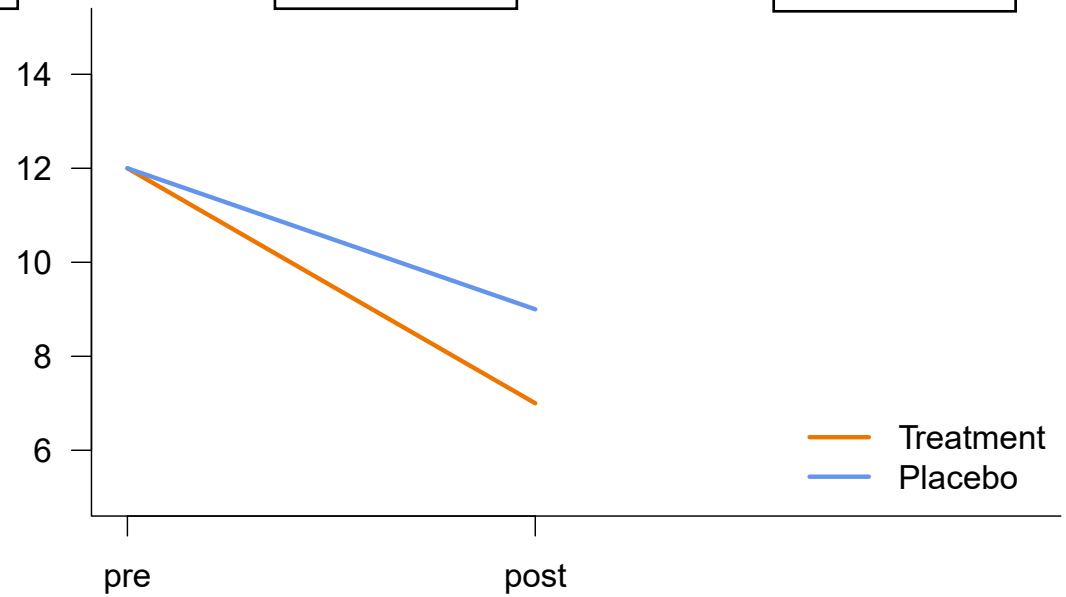
Pretest/posttest design with '*double blind* placebo control'



Photo from commons.wikimedia.org



~~Maturation/spontaneous remission~~
~~History threat~~
~~Regression to the mean~~
~~Non-specific effect (*not in book!*)~~
~~Placebo~~
~~Observer bias~~
~~Demand characteristics~~
Attrition / subject loss
~~Testing threat~~
~~Instrumentation threat~~

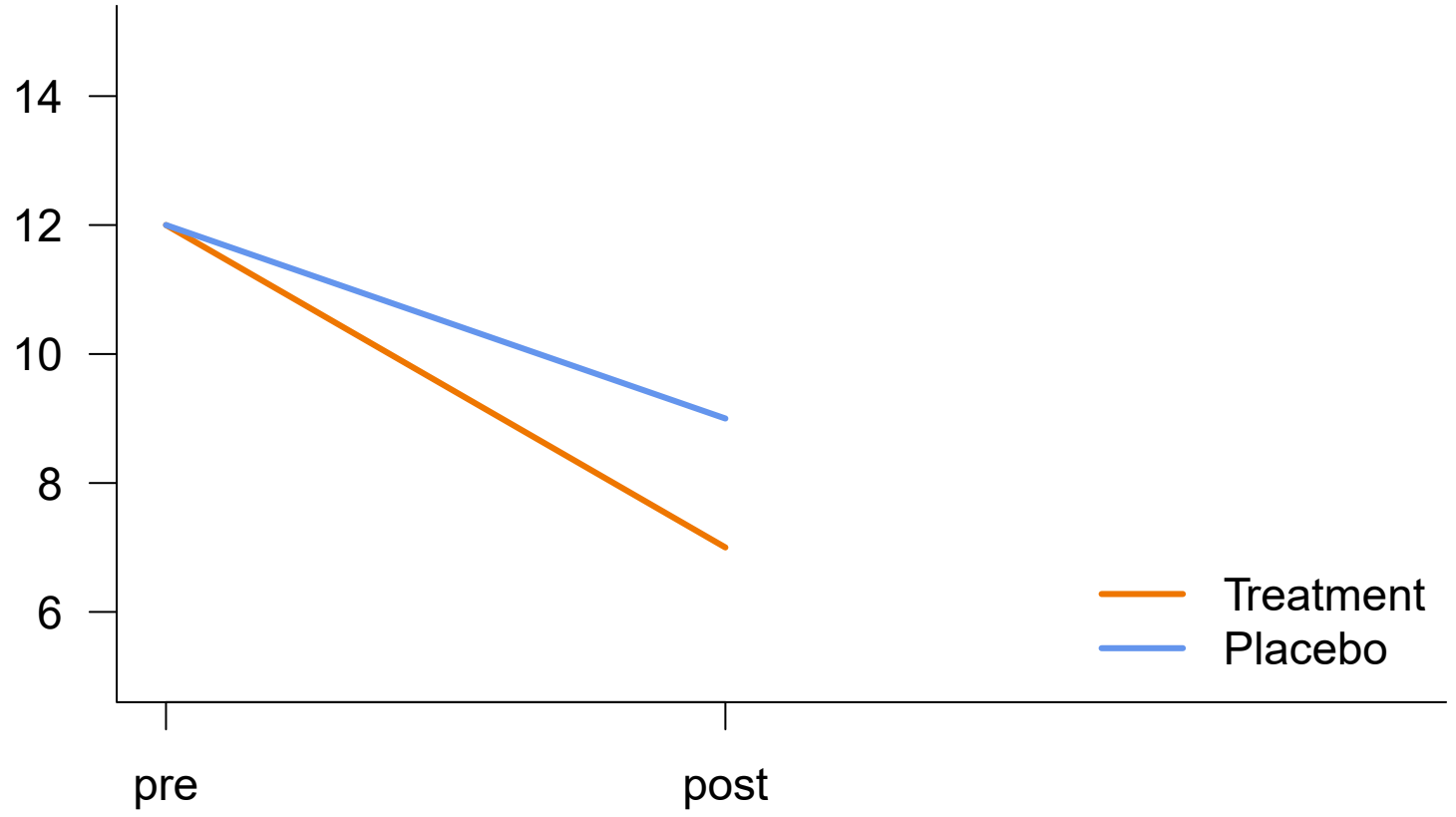


Attrition

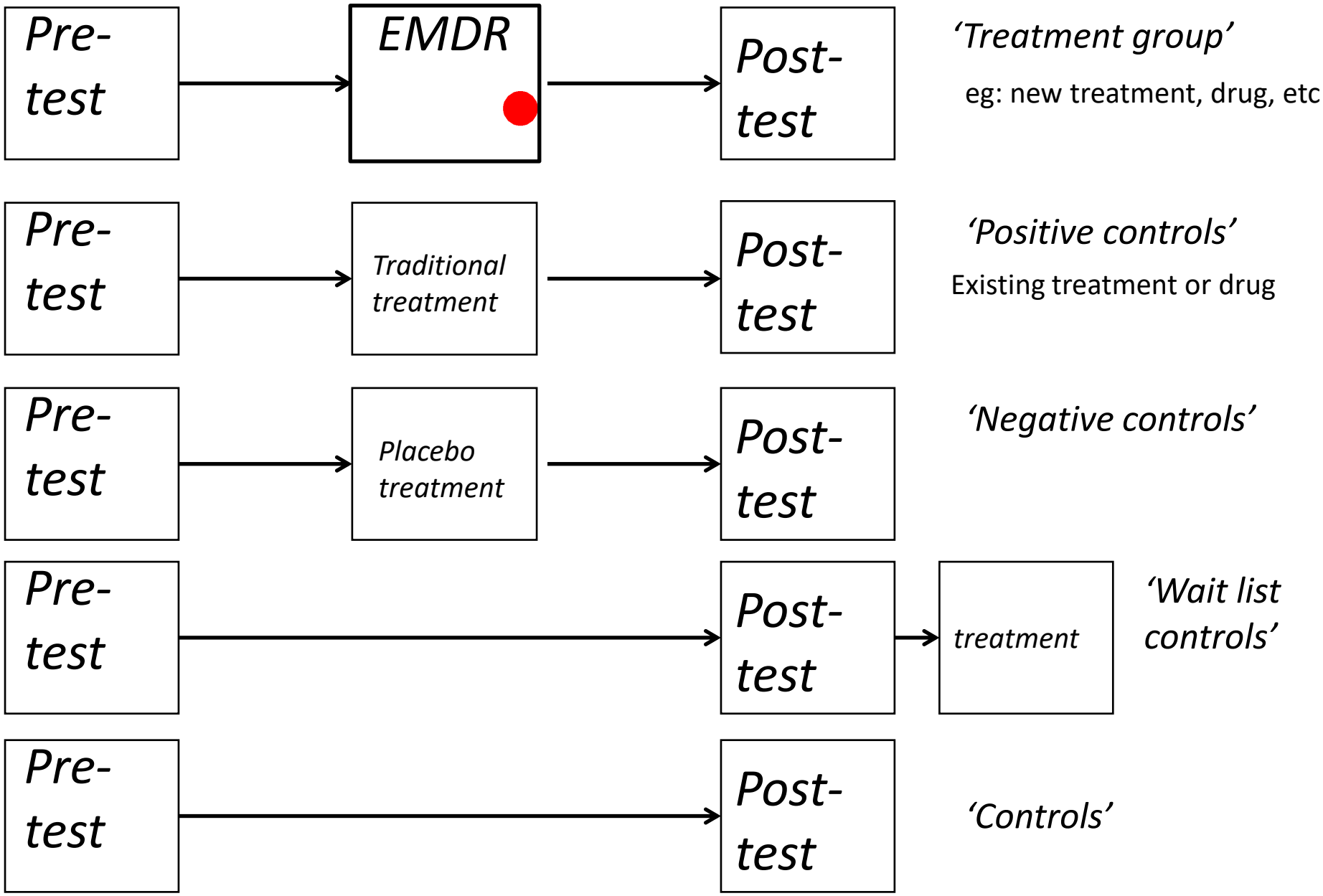
- Attrition:
  - People may drop-out
  - Relatively unproblematic if drop-out is random
  - Problematic if dropout is systematic in one condition

- What to do:
  - In advance, think about what may cause attrition
  - Easy to identify whether it’s present. So, check and anlyze.
  - For people who drop out before post test, also remove scores pretest (so remove completely)

Example: Was there a placebo effect?

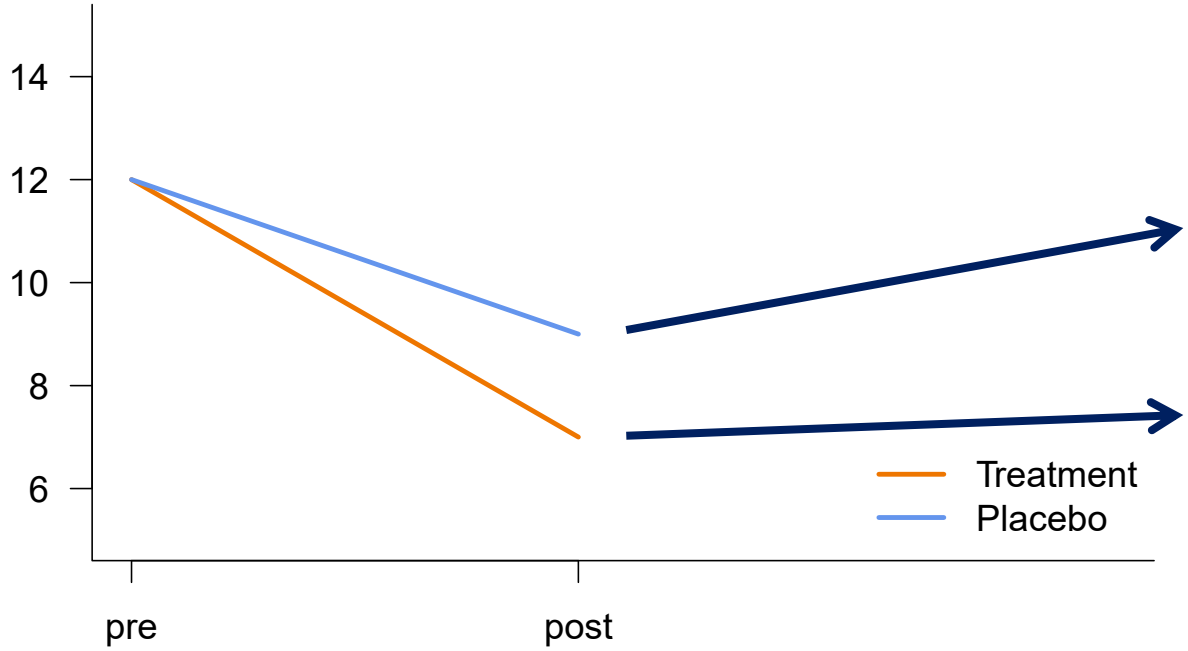


Multiple control groups



Example: Was there a treatment effect?

(Assume double-blind placebo group and no drop-out)



- Maturation/spontaneous remission
- History threat
- Regression to the mean
- Non-specific effect
- Placebo effect

\-

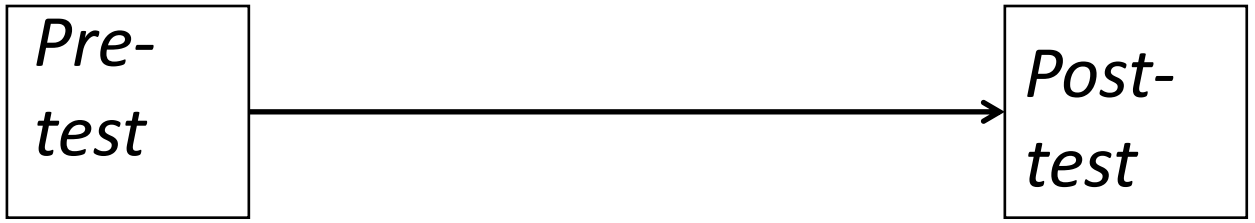
- Maturation/spontaneous remission
- History threat
- Regression to the mean
- Non-specific effect
- Placebo effect
- Treatment

Difference:

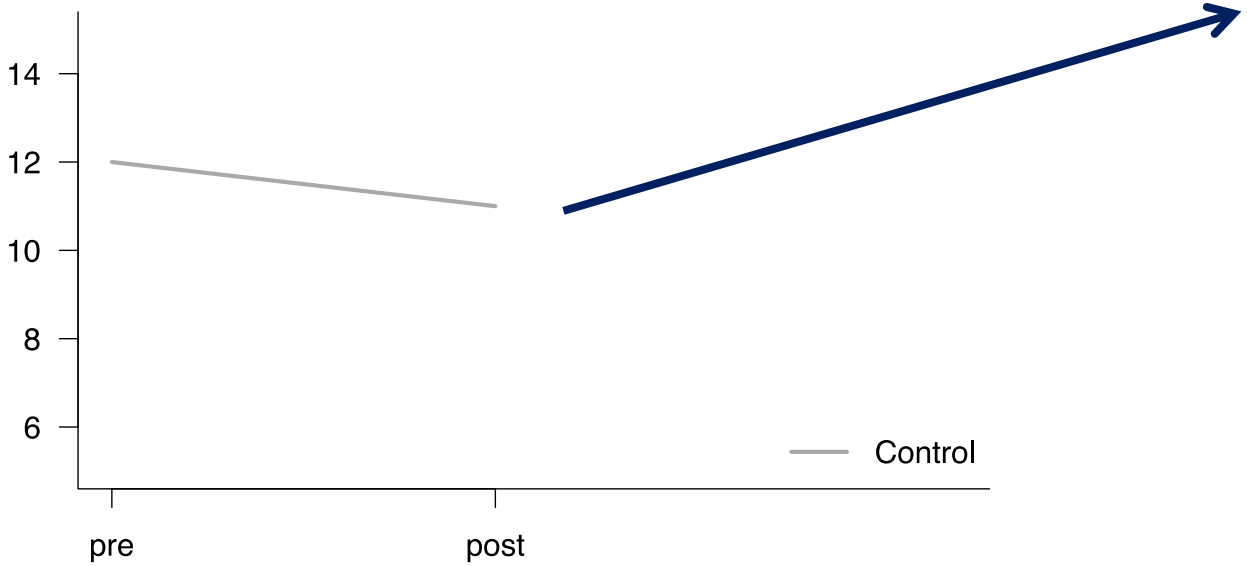
- *Treatment*

"Substract it out"

Multiple control groups

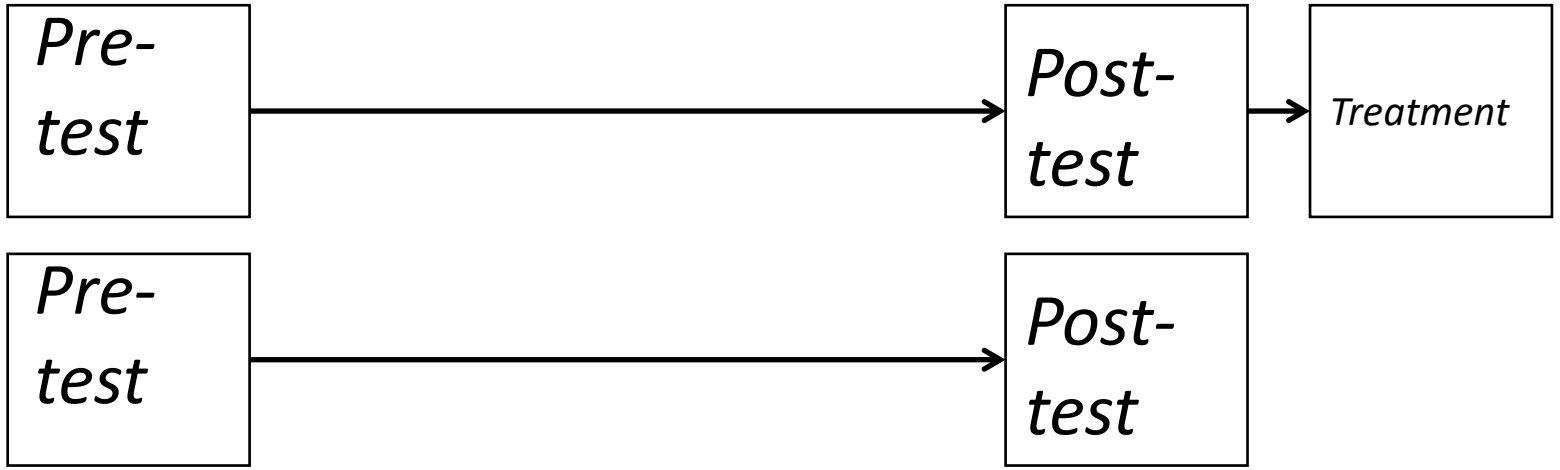


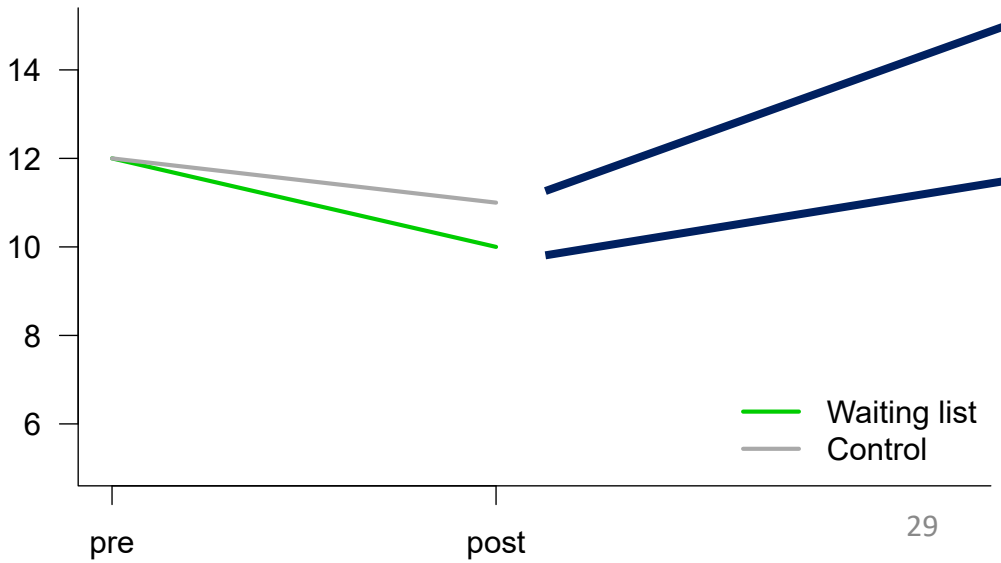
'Controls'



- Maturation/spontaneous remission
- History threat
- Regression to the mean

Multiple control groups





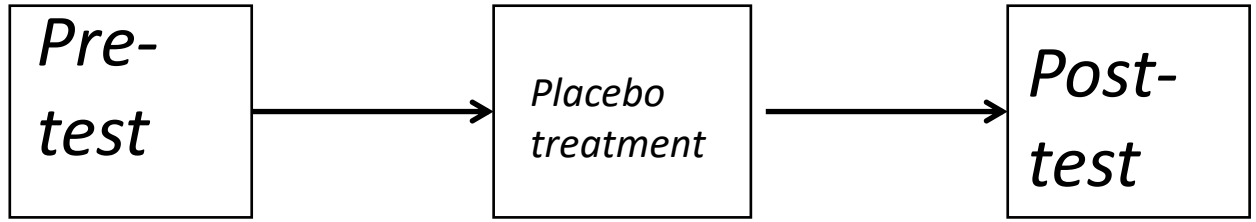
- Maturation/spontaneous remission
- History threat
- Regression to the mean

- Maturation/spontaneous remission
- History threat
- Regression to the mean
- Non-specific effects

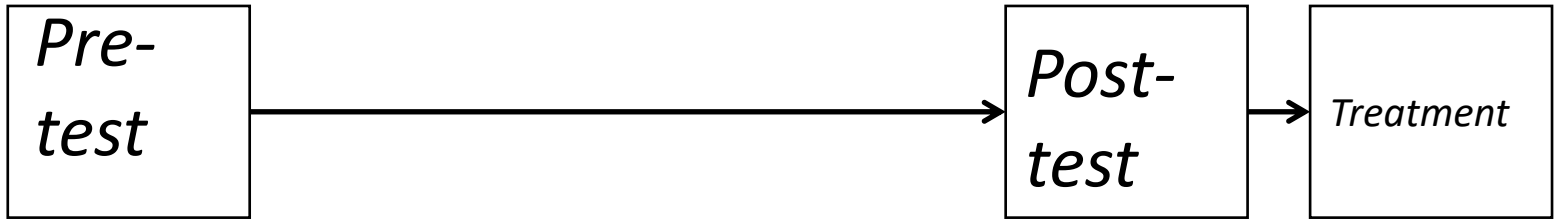
Difference:

- ***Non-specific effects***

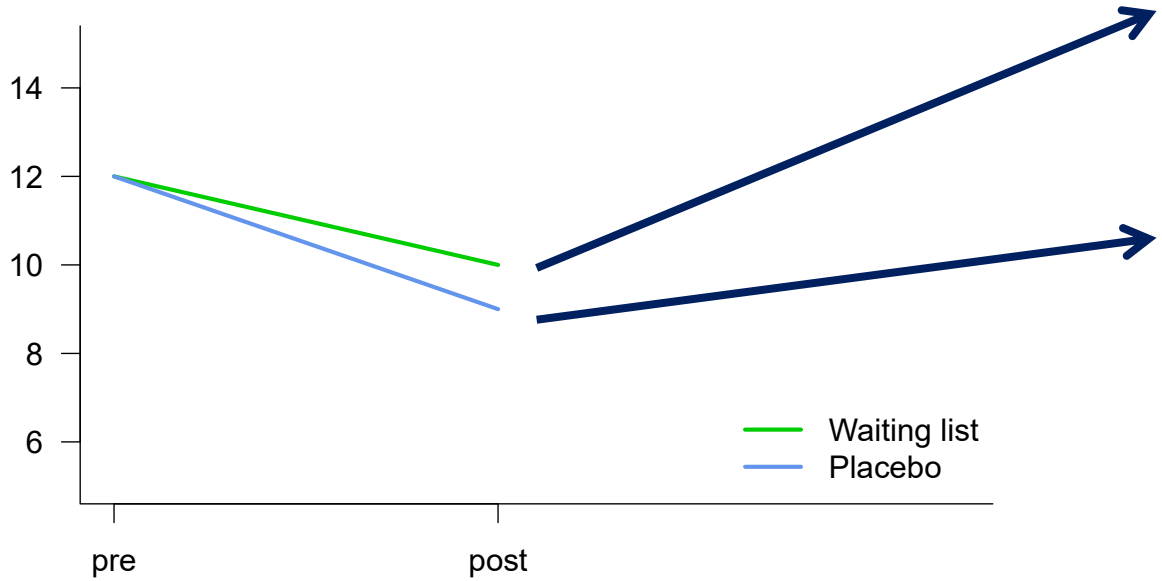
Multiple control groups



'Negative controls'



'Wait list controls'



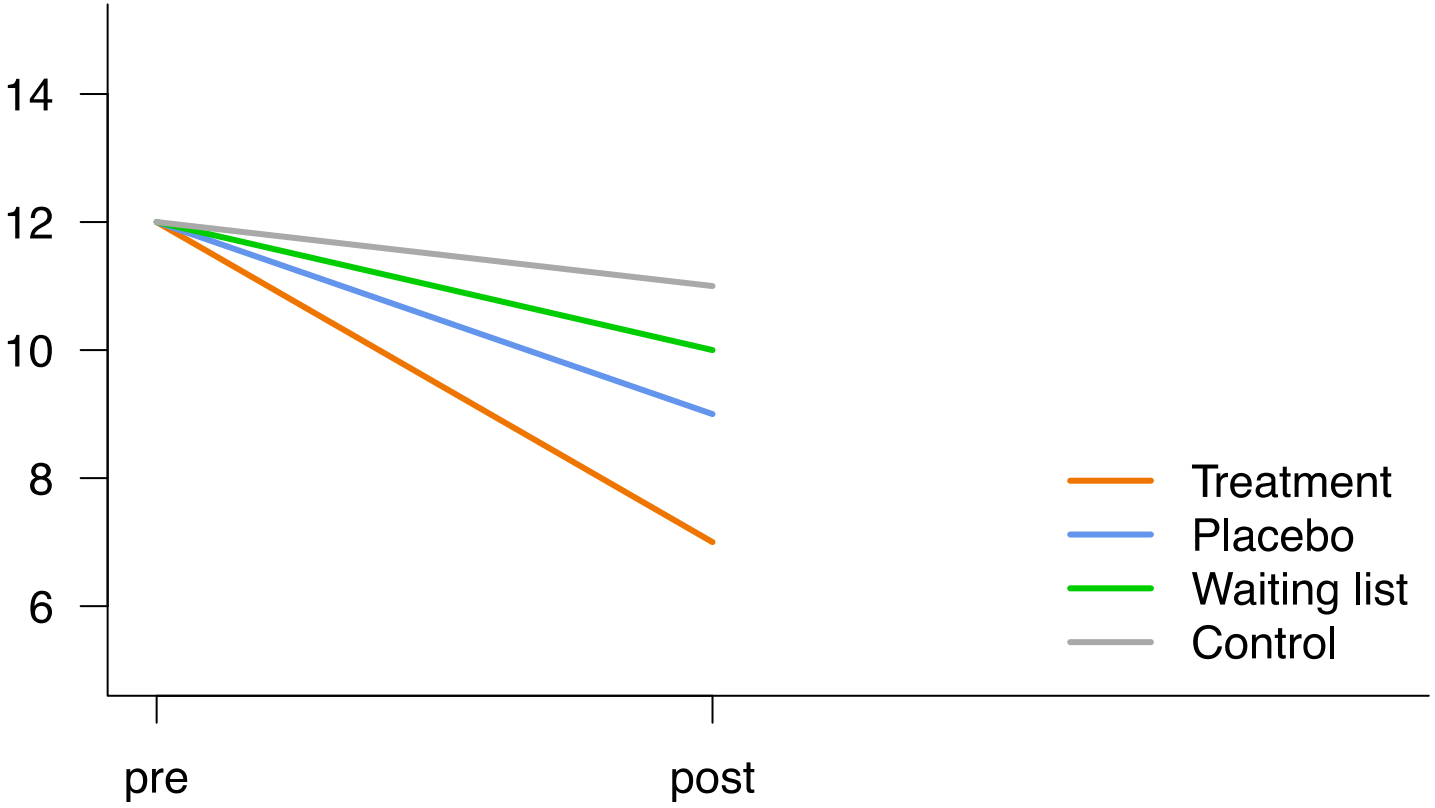
- Maturation/spontaneous remission
- History threat
- Regression to the mean
- Non-specific effects

- Maturation/spontaneous remission
- History threat
- Regression to the mean
- Non-specific effects
- Placebo effect

Difference:

- ***Placebo effect***

Including all control groups



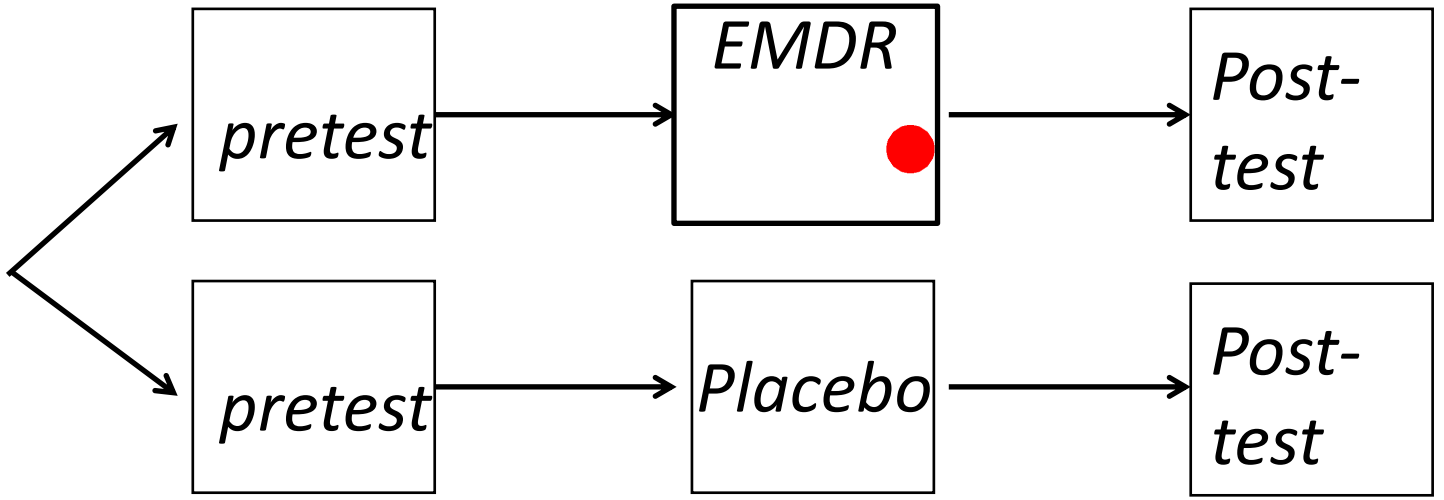
- Maturation threat
- History threat
- Regression to the mean
- Non-specific effects vs
- Placebo effect vs
- Treatment vs

Design #5: best design to test treatment effect

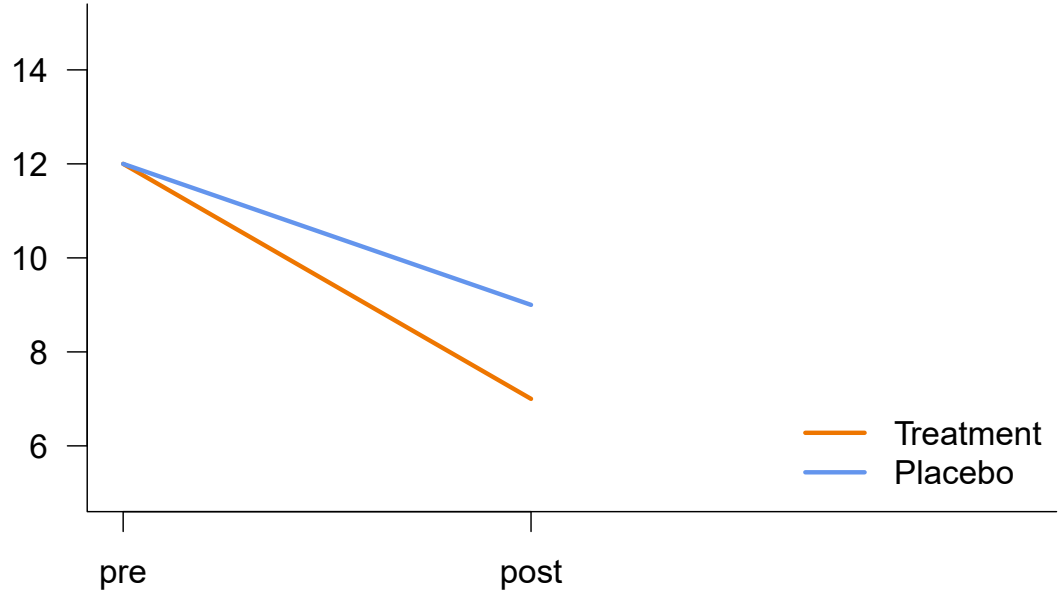
Pretest/posttest design with '*double blind* placebo control'



Photo from commons.wikimedia.org







Today

1. Large *N* clinical trials
2. **Small *N* clinical trials**
3. Quasi-experiments

Why small *N*?

- Some conditions are extremely rare, and finding a population >1 is a challenge
  - E.g.: face blindness: inability to recognize faces (including your own!)

- In clinical practice, individuals differ in their symptoms
  - In some cases you might be interested in whether something works for *this* person, not whether it works *in general*

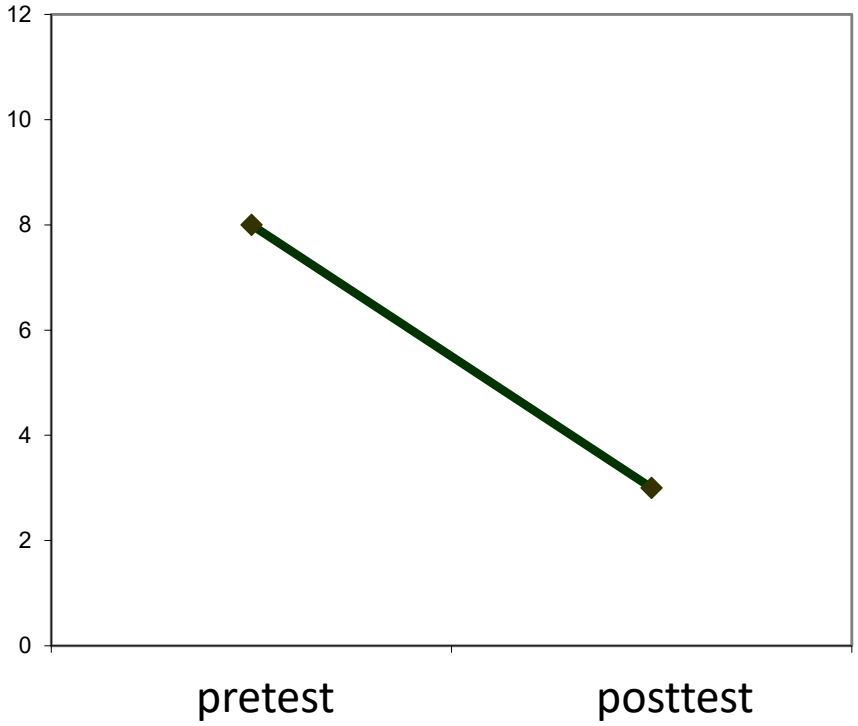
Small N vs Large N

- Large N
  - Interest in inferences about **populations** (eg., mean difference)
  - To determine the validity of the group mean, use an inferential test
    - E.g., confidence interval

- Small N
  - When you are interested in the **individual**
  - To determine the validity of an effect *within one* individual, use replication
    - Within-subjects design

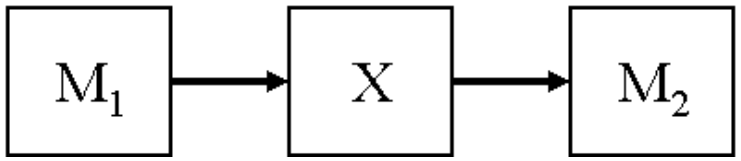
Intervention

Is the treatment effective?



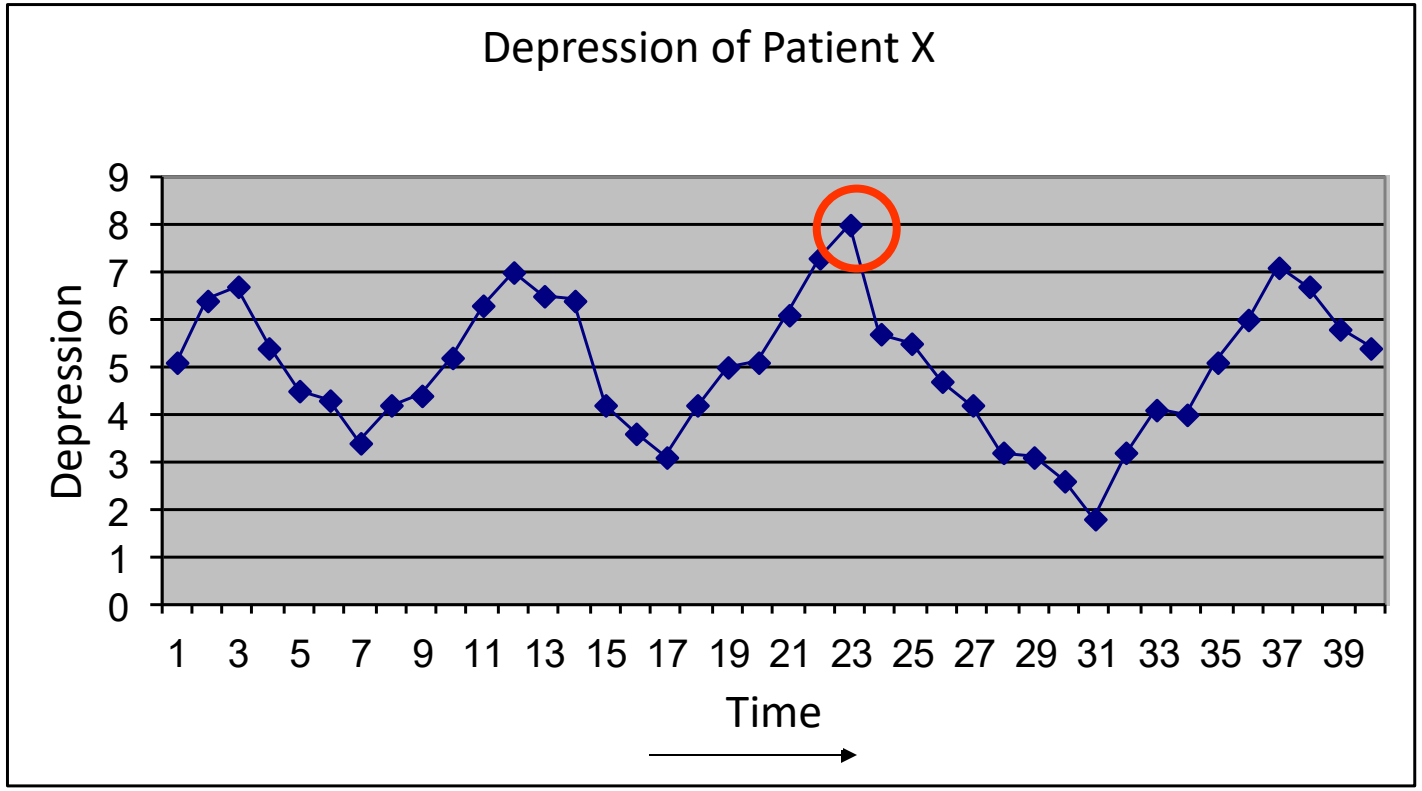
Problem: No control group!

Threat to internal validity



Example threat: Regression to the mean

Treatment started



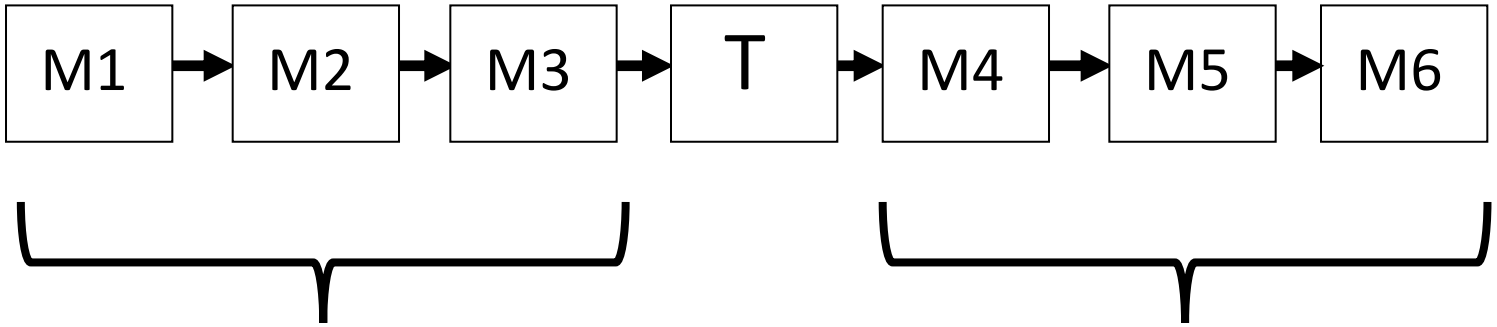
Possible small N designs

Solutions to threats to internal validity

1. Stable baseline design
2. Reversal design
3. Multiple baseline design

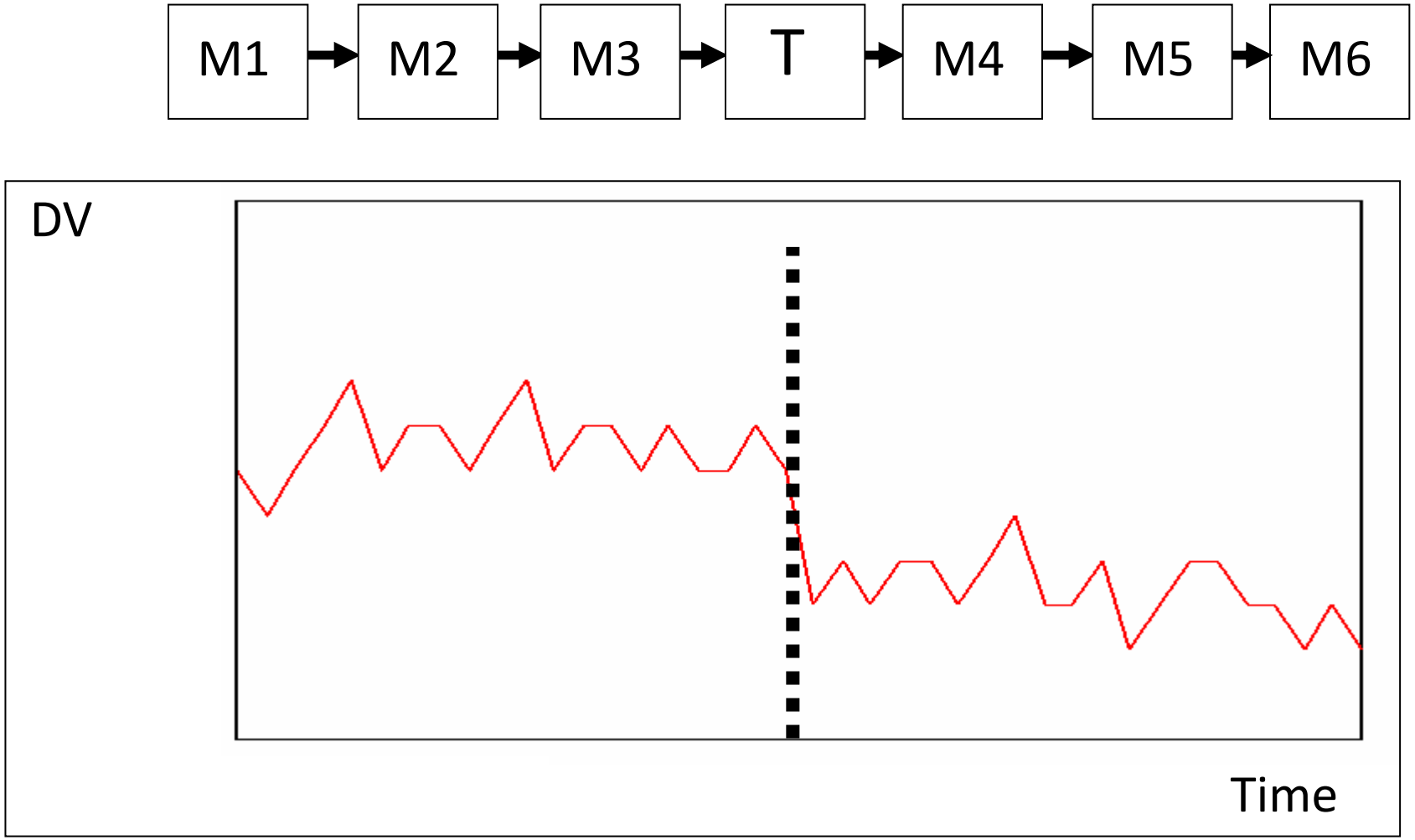
Stable baseline design

- M: Measurement
- T: Treatment



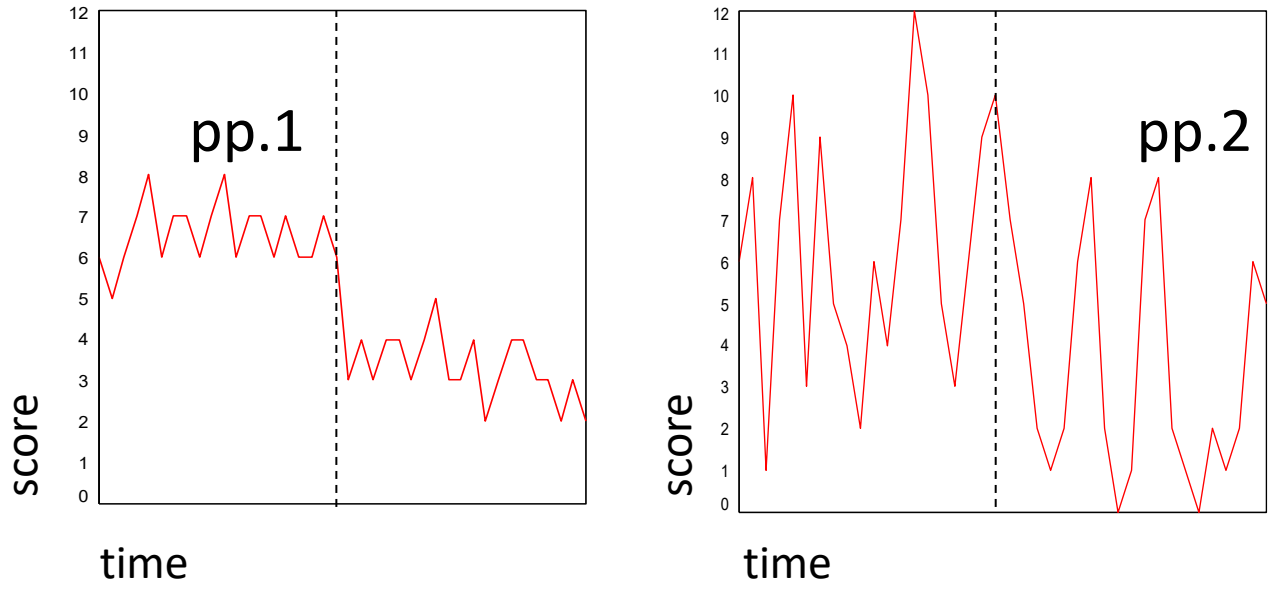
Replication to assess stability of an extended baseline period

Stable baseline design



Stable baseline design

- Controls for *some* internal validity threats
  - Maturation
  - History
  - Non-specific effect
  - Regression to the mean

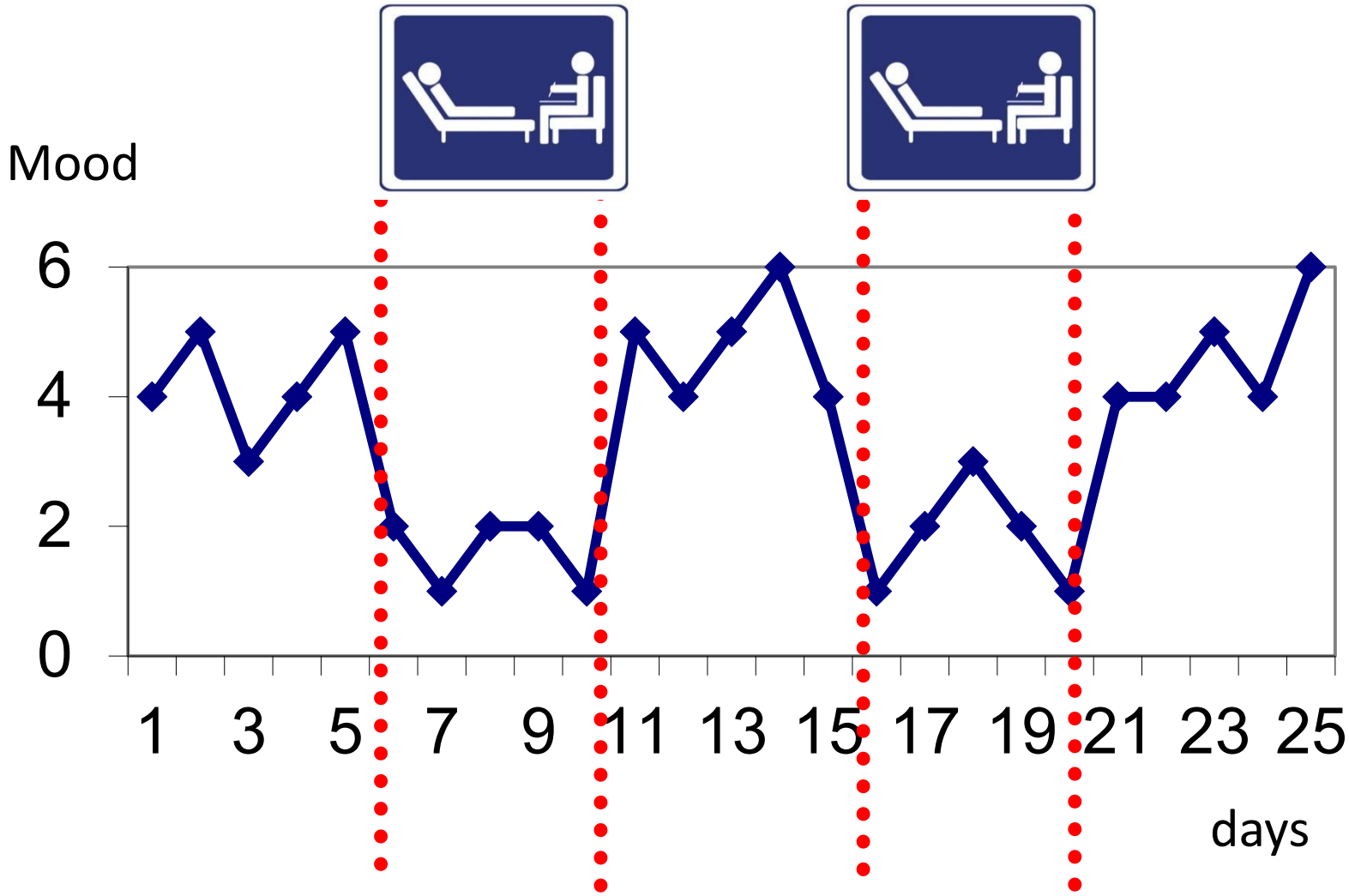


Reversal design

Researchers take the treatment away for a time (the reversal period) to see whether the problem returns.

This gives additional evidence that treatment has an effect, but this design is only appropriate in case one expects that treatment does not cause lasting change.

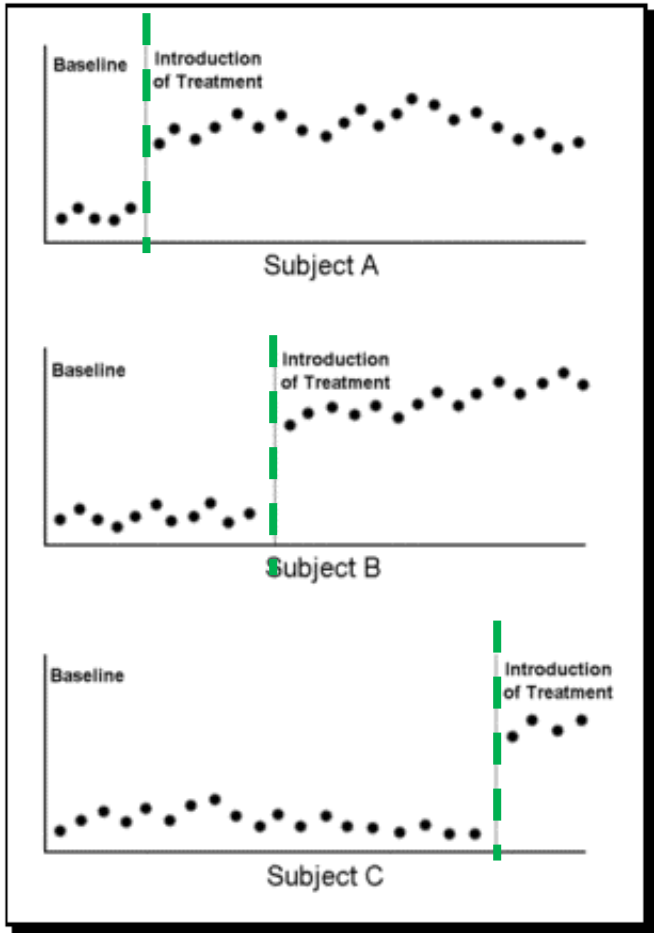
And it might also not always be ethical!



Multiple baseline design

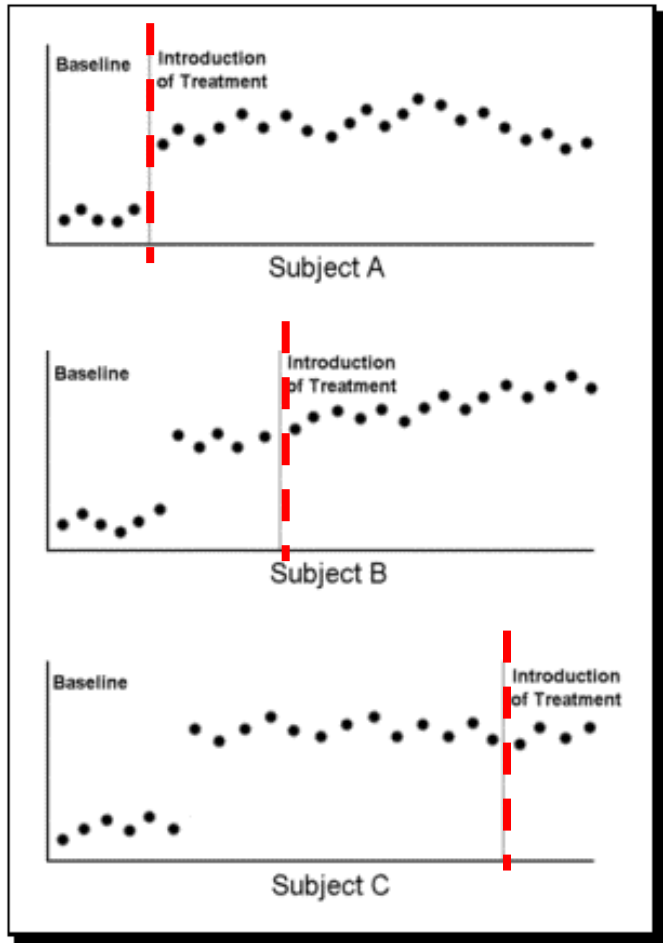
- Study effect of intervention in multiple baseline designs
  - Consider multiple individuals, multiple settings or multiple behaviors *at different times*

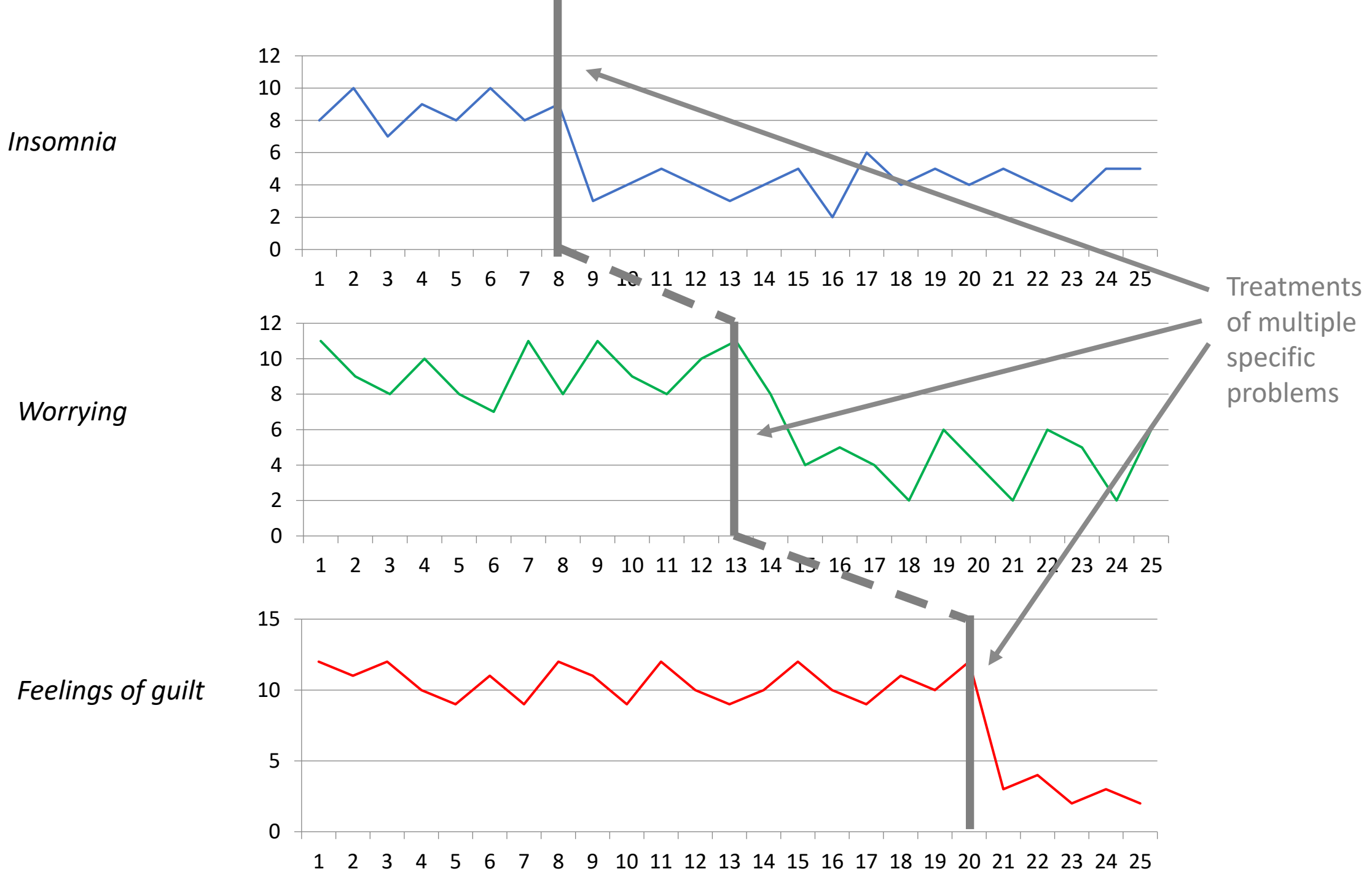
Multiple baseline design



← Successful treatment

Unsuccessful treatment →




Insomnia
Worrying
Feelings of guilt
Treatments of multiple specific problems

Today

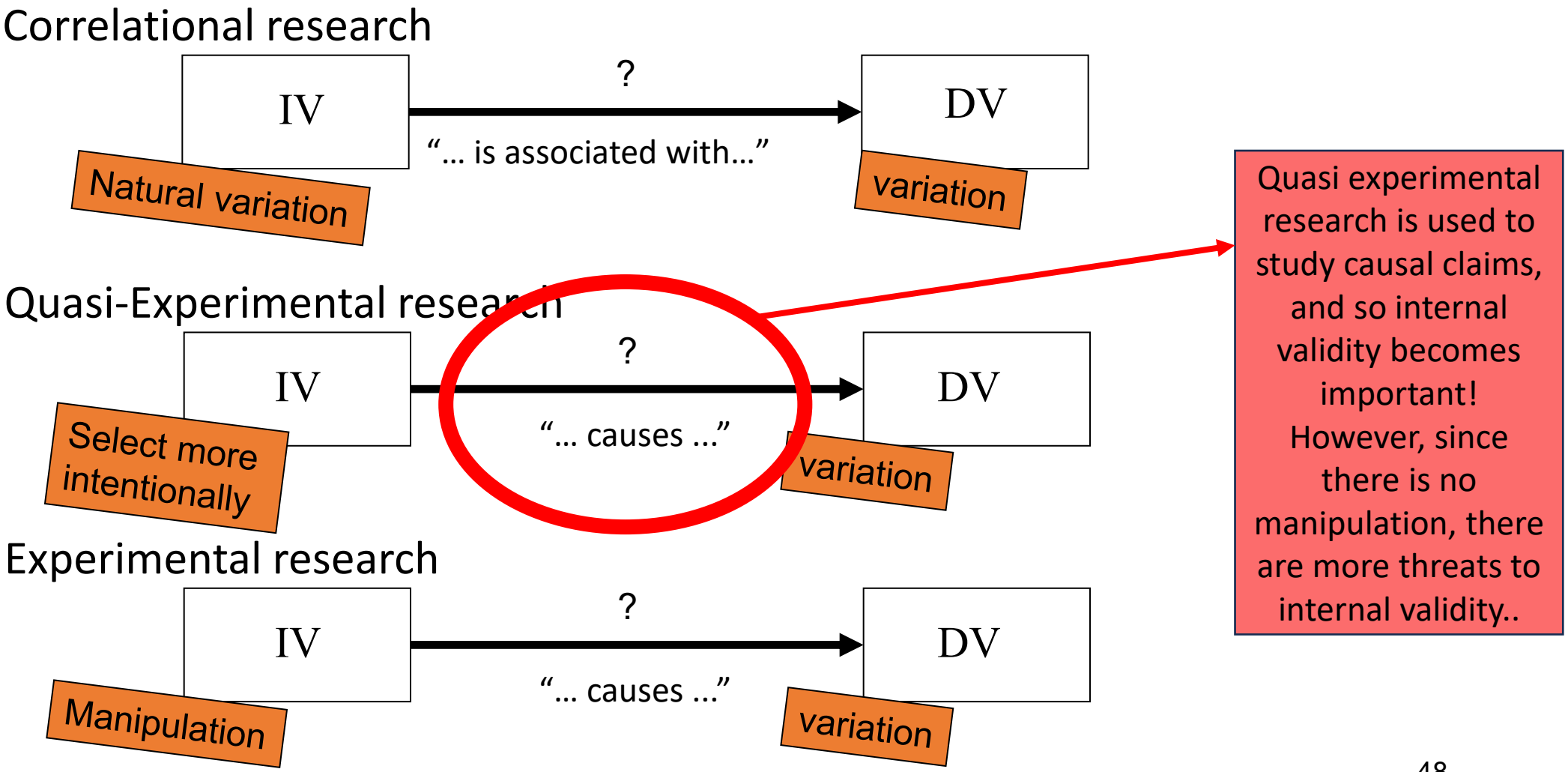
1. Large *N* clinical trials
2. Small *N* clinical trials
3. **Quasi-experiments**

Quasi-Experimental research

Sometimes experimental manipulations are impossible/undesirable

- Unethical
  - e.g., health effects of smoking
- Unpractical
  - e.g., the effect of watching Sesame Street on language development
- Impossible
  - e.g., effect of extraversion on job success

Quasi-Experimental research



Between-group quasi-experiments

Difference with real experiment is that there is *no random assignment*, the independent variable is some already scheduled event (is not in the researchers control)

- Non-equivalent control posttest design
  - Select two groups that differ on independent variable
    - E.g., the bridge study from lecture 7
  - Because there is no random assignment, there is no guarantee that the groups are equivalent
- See Ch. 13 for examples of other designs
  - Non-equivalent control pretest/posttest design
  - Interrupted time series design
  - Combi: Nonequivalent control group interrupted time-series design





Photo: https://www.geograph.org.uk/photo/6181189 Photo: needpix.com

Quasi-experimental studies

- Internal validity is always questionable
  - Due to selection effects in particular
  - But also all other threats discussed before

- There is often a trade-off between internal validity and external validity
  - real experiment: enhances internal validity
  - quasi-experiment: enhances external validity

Chapter 4 & 14

- Chapter 4 on ethics and Chapter 14 on replication are also exam material, but not covered in the lectures or WAs

- You can study these chapter by yourself, topics are not too difficult

Summary

- To study the effectiveness of a treatment or prevention, clinical psychology makes use of large-N clinical trials.
- In these trials, it is important to choose a design that rules out as many threats to internal validity as possible
- Depending on the study, different threats can ask for different study designs, but generally a mixed design with ‘double blind placebo control’ rules out many of the confounders that threaten the internal validity of a clinical trial
- Sometimes, a researcher can be interested in studying the effect of an intervention on a single individual (or a couple of individuals), which asks for small-N designs that ensure as much internal validity as possible
- Sometimes experimental intervention/manipulation is impossible or unwanted, in which case one can choose to do a *quasi experiment*
- Because there is no random assignment in quasi experiments, internal validity is harder to reach. It is therefore crucial in quasi experimental studies to think about what could be possible confounders and how to avoid them

Example question

Aarush is doing an experiment to study the effect of alcohol on reaction time. He assigns people to two conditions. In the alcohol condition, participants get to drink cola with alcohol while in the control condition participants get to drink cola without alcohol. To mask the taste of alcohol he adds a drop of peppermint oil to the drinks in both conditions. The peppermint oil should avoid:

a) Attrition
b) Demand characteristics
c) Instrumentation threat

Example question

Aarush is doing an experiment to study the effect of alcohol on reaction time. He assigns people to two conditions. In the alcohol condition, participants get to drink cola with alcohol while in the control condition participants get to drink cola without alcohol. To mask the taste of alcohol he adds a drop of peppermint oil to the drinks in both conditions. The peppermint oil should avoid :

a) Attrition

b) Demand characteristics

c) Instrumentation threat