THE UNIVERSITY OF MELBOURNE

Faculty of Medicine, Dentistry and Health Sciences

MACH Melbourne Academic Centre for Health

Analysing change from baseline in trials: what is the best approach?

Ms Fiona McManus

✉ fmcmanus@unimelb.edu.au

Methods and Implementation Support for Clinical and Health research Hub

Centre for Epidemiology and Biostatistics

Melbourne School of Population and Global Health

Website:- https://clinicalresearch.mdhs.unimelb.edu.au/

Email:- misch-info@unimelb.edu.au @MischHub

Housekeeping

- Please keep your microphone switched off during the presentation.
- You are welcome to leave your video on or off as you prefer.
- If you have any questions, please feel free to enter them in the chat box. We will review them throughout the presentation.
- Note that this presentation will be recorded and a link will be provided after the webinar.
- A copy of the slides will also be provided.

MISCH Hub

- **M**ethods and **I**mplementation **S**upport for **C**linical and **H**ealth research Hub

- Our aim is to provide support to researchers and affiliated researchers of the University of Melbourne in health research.

- We provide support on core research methods of Biostatistics and Clinical Epidemiology, Health Economics, Clinical Trials, Implementation Effectiveness and Co-Design and Health Informatics (REDCap).

MISCH Staff


BIOSTATISTICIAN
Fiona McManus


BIOSTATISTICIAN
Patty Chondros


BIOSTATISTICIAN
Emily Karahalios


HEALTH ECONOMIST
Kim Dalziel


EPIDEMIOLOGIST
Mark Jenkins


CO-DESIGN AND IMPLEMENTATION EFFECTIVENESS LEAD
Cathy Vaughan


CLINICAL TRIALS LEAD
Adam Deane


MISCH MANAGER
Shale O'Brien


CO-LEAD, MISCH BIOSTATISTICS
Sabine Braat


MISCH OPERATIONS MANAGER
An Nguyen


BIOSTATISTICS LEAD, AUSTIN HOSPITAL
Leonid Churilov


BIOSTATISTICIAN
Dominic Italiano


DIRECTOR OF MISCH
Julie Simpson


BIOSTATISTICIAN
Rob Mahar


EPIDEMIOLOGIST AND BIOSTATISTICIAN
Digsu Koye


HEALTH ECONOMIST
Paul Amores


HEALTH ECONOMIST LEAD
An Duy Tran


BIOSTATISTICIAN
Alistair McLean


BIOSTATISTICIAN
Anurika De Silva


BIOSTATISTICIAN
Hannah Johns


HEALTH INFORMATICS LEAD
Douglas Boyle


BIOSTATISTICIAN
Joanna Ling


CO-LEAD, MISCH BIOSTATISTICS
Karen Lamb


BIOSTATISTICIAN
Sophie Zaloumis


BIOSTATISTICIAN
Vanessa Pac Soo


CLINICAL TRIALS OFFICER
Katie Ozdowska


HEALTH INFORMATICIAN
David Ormiston-Smith


CO-DESIGN AND IMPLEMENTATION EFFECTIVENESS
Erin Davis

Analysis of change

- Researchers commonly obtain measurements on participants at the beginning of the study (**baseline**) and then at some time point after an intervention has been applied (**post-intervention**).

Table 1. Study schedule

Evaluation	Screening	Entry	Post-Entry Evaluations (Weeks)							
			4	12	24	36	48	60	72	Discontinuation Evaluations
			± 7 days		± 14 days					
Medical/Medication History	X									
Clinical Assessments	X	X	X	X	X	X	X	X	X	X
Quality of Life **Questionnaires**		X					X			X
Demographics Questionnaire		**X**								
Adherence **Questionnaires**		**X**	X	X	X	X	X	X	X	X
Health Literacy Questionnaires		**X**								
Pill Count			X	X	X	X	X	X	X	X

Analysis of change

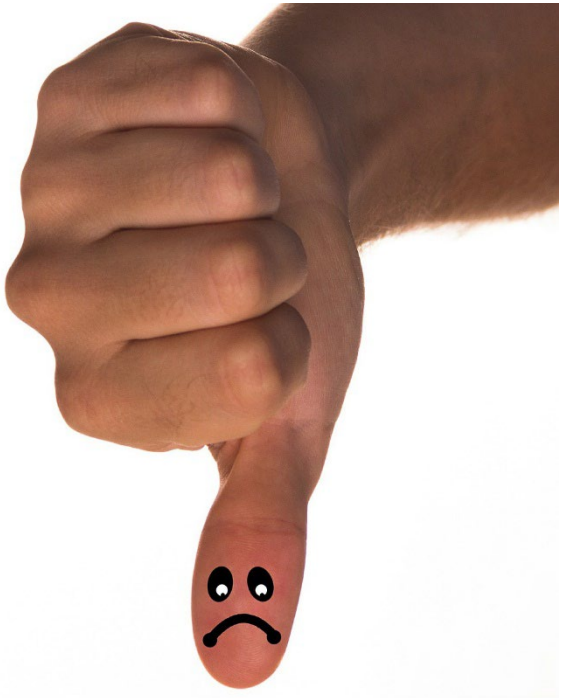
- After collecting this baseline information, we want to use these measurements in assessing the intervention effect.
- There are many possible analysis options so how do we choose the best one?

Analysis of change – problematic common approaches

- Differences: Outcome post-intervention **minus** outcome at baseline
- Percentage change:

  (Outcome post-intervention **minus** baseline)

  divided by baseline
- These approaches reduce the follow-up and baseline measurements to a **single value**.
- This makes them appealing for simple analysis (e.g., t-test).



Analysis of change – the best approach

- The best way of using baseline measurements is to **condition on baseline values** such as in a linear regression model by including them as a covariate in the model.
  - This type of approach is known as an analysis of covariance (ANCOVA).

- **Conditioning on baseline values**
  - Often referred to as ‘adjusting for baseline’ or ‘controlling for baseline’
  - Helps us assess the effect of treatment at follow-up among individuals who have the same baseline value.

Analysis of change – the best approach

- Although ‘adjusting for baseline’ is a recommended approach, many other methods are still used.
- The aim of this presentation is to highlight why ‘adjusting for baseline’ is the recommended approach and why others are less suitable.

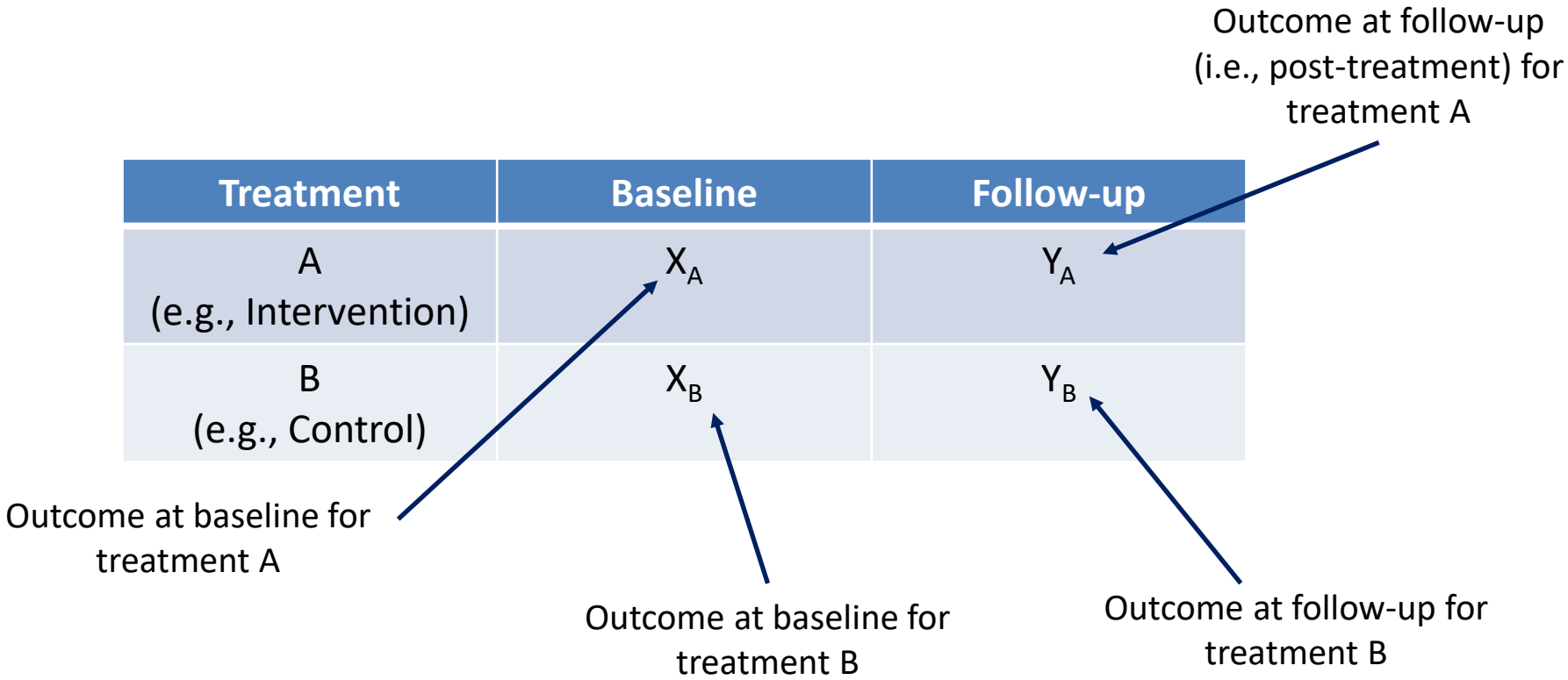
Handling baseline measurements in two-group comparisons

- Consider a clinical trial of an exercise program to reduce pain in people with knee osteoarthritis.
- Control group receives education only for 6 weeks.
- Intervention arm receives education and an exercise program for 6 weeks.
- Pain is measured before and after treatment on a numerical rating scale (NRS).

Treatment	Time 1 (Baseline)	Time 2 (Follow-up)
Exercise (Intervention)	Pain (NRS)	Pain (NRS)
Education (Control)	Pain (NRS)	Pain (NRS)

Mathematical notation to follow!

Handling baseline measurements in two-group comparisons



Handling baseline measurements in two-group comparisons

Treatment	Baseline	Follow-up
A (e.g., Intervention)	X_A	Y_A
B (e.g., Control)	X_B	Y_B

- $\overline{X_A}$ is the baseline mean for treatment A (e.g., the mean pain score before treatment for those who received exercise).

- $\overline{X_B}$ is the baseline mean for treatment B (e.g., the mean pain score before treatment for those who received education only).

Handling baseline measurements in two-group comparisons

Treatment	Baseline	Follow-up
A (e.g., Intervention)	X_A	Y_A
B (e.g., Control)	X_B	Y_B

- $\overline{Y_A}$ is the post-treatment mean for treatment A (e.g., the mean pain score after treatment for those who received exercise).

- $\overline{Y_B}$ is the post-treatment mean for treatment B (e.g., the mean pain score after treatment for those who received education only).

Handling baseline measurements in two-group comparisons

Treatment	Baseline	Follow-up
A (e.g., Intervention)	X_A	Y_A
B (e.g., Control)	X_B	Y_B

- Interested in comparing the groups
  - E.g., comparing the mean outcome after treatment between the groups (i.e., $\overline{Y_A} - \overline{Y_B}$).

- You notice that the distribution of the baseline measurements differs between the two groups (i.e., there is some imbalance in the outcome at baseline).
- That is, the mean baseline pain score for those who received exercise (i.e., $\overline{X_A}$) does not equal the mean baseline pain score for those who received education (i.e., $\overline{X_B}$).
- Want to ensure baseline values (i.e., X_A , X_B) are accounted for.

Handling baseline measurements in two-group comparisons

Here are three possible methods:

1) Estimate the between-group **difference in means for post-intervention outcomes** only (i.e., ignore baseline values altogether).

2) Estimate the between-group **difference in means for post-intervention minus baseline values** (e.g., derive the change in outcome for each participant).

3) Estimate the between-group **difference in means for either i) post-intervention outcomes adjusted for baseline or ii) change in outcomes (e.g. post – baseline) adjusted for baseline** (i.e. condition on baseline values).
   - E.g., using analysis of covariance (ANCOVA) or multivariable linear regression.

Vickers & Altman (2001) BMJ 323: 1123

Method 1: Difference in post-intervention means

- This approach ignores the baseline measurements altogether. The treatment effect (or between-group difference) is estimated as $\overline{Y_A} - \overline{Y_B}$.
- This approach assumes the **baseline measurements** are **balanced**.
- In other words, it is assumed that the baseline mean for treatment A is the same as the baseline mean for treatment B (i.e., $\overline{X_A} - \overline{X_B} = 0$).
- Ignoring values of the outcome at baseline can lead to an over- or under-estimation of the treatment effect.

Difference in post-intervention means

- Comparing post-intervention means when there is baseline imbalance results in **biased** findings.
- **'Biased'** means the mean treatment effect is **over- or under-estimated**.
- Considering only the post-intervention outcomes is also **inefficient**.
- This means it generally has **lower power** so you need a larger sample size to detect differences in post-intervention outcomes if baseline and follow-up scores are correlated.

Method 2: Difference in means for change (post-intervention minus baseline)

- This approach completely takes into account the baseline measurement in that it becomes part of the outcome. The treatment effect is estimated as $(\overline{Y_A} - \overline{X_A})$ - $(\overline{Y_B} - \overline{X_B})$ = $(\overline{Y_A} - \overline{Y_B})$ - $(\overline{X_A} - \overline{X_B})$

- However, it does not take into account the actual **correlation** between the baseline and post-intervention scores.

- Ignoring the correlation will **attenuate** the estimate of the true treatment effect.

Method 1 and Method 2

- Both method 1 and method 2, neither of which adjust for baseline, result in **biased estimates** of the treatment effect when baseline imbalance exists.
- Need to find an estimator that is **unbiased** and appropriately takes into account the correlation between baseline and post-intervention measurements.

Method 3: Adjust for baseline values

- When the outcome is post-intervention scores, the treatment effect, adjusted for baseline, is estimated as $(\overline{Y_A} - \overline{Y_B}) - \boldsymbol{\rho}(\overline{X_A} - \overline{X_B})$.
- We can adjust for baseline values of the outcome using ANCOVA or multivariable linear regression.
- This is an **unbiased** method to account for baseline values.
- It is also generally the **most efficient**.
  - More efficient = narrower confidence intervals

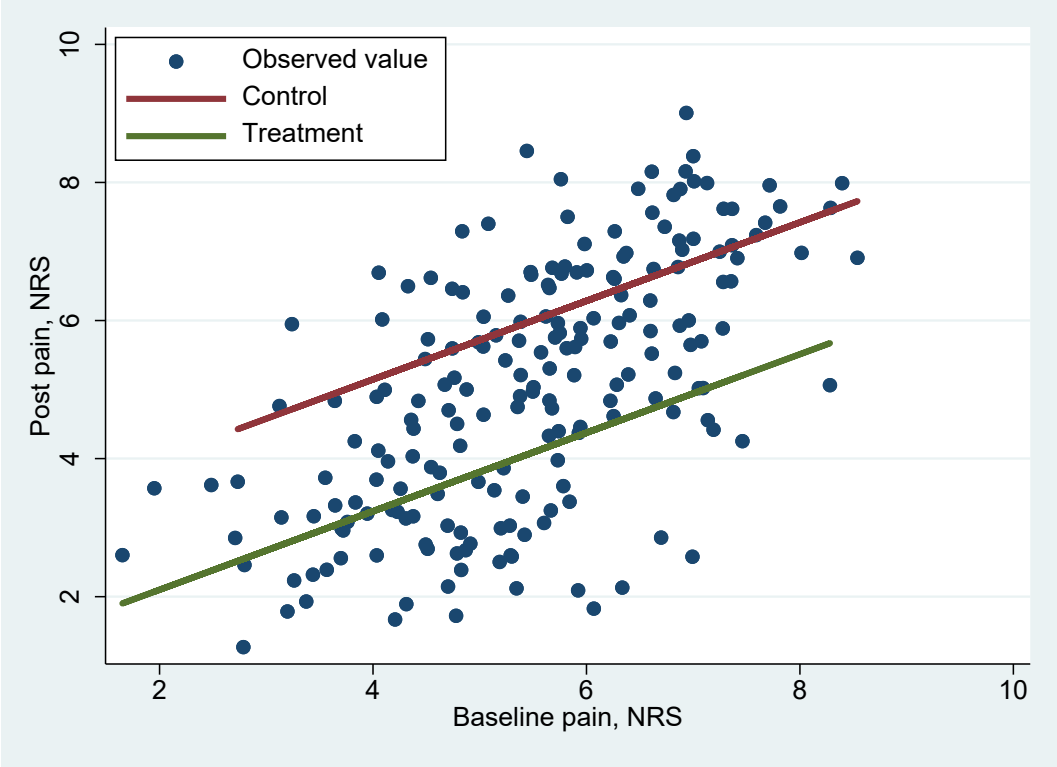
Correlation between baseline and post-intervention measures

Simulated example of pain

	Exercise (Intervention/ Treatment A) N = 100 Mean (SD)	Education (Control/ Treatment B) N = 100 Mean (SD)	Mean difference (Intervention – Control)
Baseline pain (NRS)	5.0 (1.3)	6.0 (1.2)	-1.0
Post pain (NRS)	3.8 (1.4)	6.3 (1.3)	-2.5
Post *minus* baseline pain (NRS)	-1.2 (1.3)	0.3 (1.1)	-1.5

- Difference in mean pain at baseline ($\overline{X_A} - \overline{X_B}$) is **-1.0**.
- Difference in mean pain at follow-up ($\overline{Y_A} - \overline{Y_B}$) is **-2.5**.
- Difference in mean change in pain ($\overline{Y_A} - \overline{X_A}$) - ($\overline{Y_B} - \overline{X_B}$) is **-1.5**.

Simulated example of pain



Positive correlation between baseline and post pain scores

Simulated example of pain: method 1

- Estimate difference in means for **post-intervention measurements** only:

```
regress post treatment

------------------------------------------------------------------------------
        post |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |  -2.474202   .1871793   -13.22   0.000    -2.843323   -2.105081
       _cons |    6.26208   .1323558    47.31   0.000     6.001072    6.523088
------------------------------------------------------------------------------
```

Note: -2.5 = ($\overline{Y_A} - \overline{Y_B}$)

Simulated example of pain: method 2

- Estimate difference in means for **change (post – baseline)**:

```
regress post_baseline treatment
------------------------------------------------------------------------------
post_basel~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |  -1.482263   .1754506    -8.45   0.000    -1.828255   -1.136271
       _cons |   .3016277   .1240623     2.43   0.016     .0569746    .5462808
------------------------------------------------------------------------------
```

Note: -1.5 = -2.5-(-1.0) = $(\overline{Y_A} - \overline{Y_B}) - (\overline{X_A} - \overline{X_B})$

Simulated example of pain: method 3

- Estimate difference in means for **post-intervention outcomes adjusting for baseline**:

Correlation between baseline and follow-up scores (ρ_1)

```
regress post baseline treatment
------------------------------------------------------------------------------
        post |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    baseline |   .5686661   .0640396     8.88   0.000     .4423749    .6949574
   treatment |  -1.91012    .170831    -11.18   0.000    -2.247012   -1.573228
       _cons |   2.872573   .3978353     7.22   0.000      2.08801    3.657136
------------------------------------------------------------------------------
```

Note: -1.9 = -2.5 - 0.6 x -1

$$= (\overline{Y_A} - \overline{Y_B}) - \rho_1(\overline{X_A} - \overline{X_B})$$

Simulated example of pain: method 3

- Estimate difference in means for **change in outcomes (post – baseline) adjusting for baseline**:

Correlation between baseline & change scores (ρ_2) = Correlation between baseline & follow up scores - 1

```
regress post_baseline baseline treatment
------------------------------------------------------------------------------
post_basel~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    baseline |  -.4313339   .0640396    -6.74   0.000    -.5576251   -.3050426
   treatment |   -1.91012    .170831   -11.18   0.000    -2.247012   -1.573228
       _cons |   2.872573   .3978353     7.22   0.000      2.08801    3.657136
------------------------------------------------------------------------------
```

Note: -1.9 = -2.5-(-1.0) – (-0.4 x -1.0) = -1.5 – 0.4

$$= (\overline{Y_A} - \overline{Y_B}) - (\overline{X_A} - \overline{X_B}) - \rho_2(\overline{X_A} - \overline{X_B})$$

$$= (\overline{Y_A} - \overline{Y_B}) - (1 + \rho_2)(\overline{X_A} - \overline{X_B})$$

Simulated example of pain: method 3 comparisons

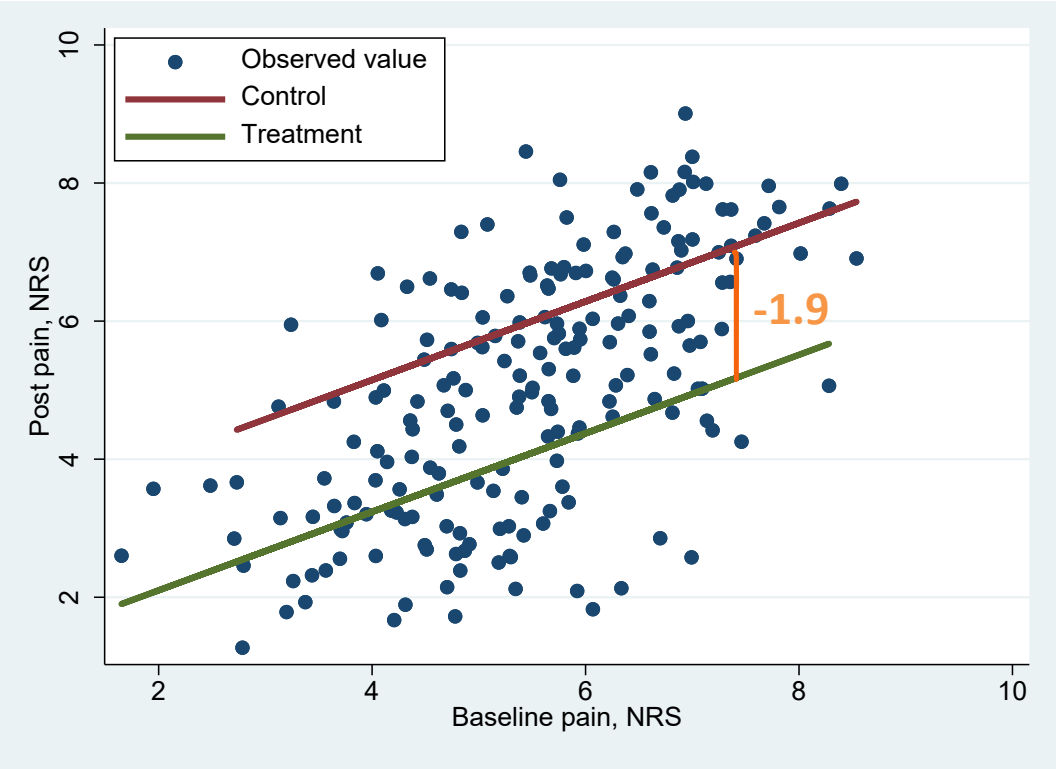
	Correlation between baseline and outcome	Mean difference (Intervention – Control) [95% confidence interval]
Post pain, adjusted for baseline (Method 3)	0.57	-1.91 [-2.25, -1.57]
Change in (post – baseline) pain, adjusted for baseline (Method 3)	-0.43	-1.91 [-2.25, -1.57]

- Between-group mean difference and 95% confidence interval are exactly the same, irrespective of whether post or change is the outcome!

Simulated example of pain: comparisons

	Exercise (Intervention/ Treatment A) N = 100	Education (Control/ Treatment B) N = 100	Mean difference (Intervention – Control)	Confidence interval width (standard error x 2)
Baseline pain (NRS)	5.0 (1.3)	6.0 (1.2)	-1.0	-
Post pain (NRS) (Method 1)	3.8 (1.4)	6.3 (1.3)	-2.5 (overestimates)	0.37
Post *minus* baseline pain (NRS) (Method 2)	-1.2 (1.3)	0.3 (1.1)	-1.5 (underestimates)	0.35
Post *or* change in pain, adjusted for baseline (Method 3)	-	-	-1.9 (unbiased)	0.34

Handling baseline in two-group comparisons



- Adjustment for baseline values compares groups while **holding the baseline values constant**

- Interpretation:

*After adjusting for baseline differences in pain, mean pain **decreased by 1.9** NRS units in those receiving exercise compared with those receiving education.*

Assumptions of adjusting for baseline

- Assumes **linear relationship** between baseline and follow-up values.
- Assumes **no interaction** between baseline outcome measurement and treatment.
  - This means it assumes that the relationship between baseline and post-intervention outcomes are the same for each intervention (i.e., treatment and control).
    - Assumes baseline outcomes do not modify the association between treatment and post-intervention outcomes.

Additional benefit of adjusting for baseline

- Adjusting for baseline generally has **greater statistical power** than the other two approaches.

- Assuming a correlation of 0.4 between baseline and follow-up pain scores, a clinically important difference of 1.8 NRS units, a standard deviation of 3 NRS units, power of 80% and significance level of 5%, the following sample sizes are required:

  - **Follow-up scores (Method 1):** a sample size of **59** per group is required.
  - **Change scores (Method 2):** a sample size of **71** per group is required.
  - **Adjusting for baseline (Method 3):** a sample size of only **50** per group is required.

In Stata: `sampsi 0 1.8, sd(3) pre(1) post(1) r01(0.4)`

What about % change?

- % Change:

  (Outcome post-intervention **minus** baseline) **divided** by baseline

- The numerator is simply the calculated change between baseline and post-intervention outcome scores.

- This means it suffers from the same problem as discussed for change from baseline when we do not adjust for baseline.

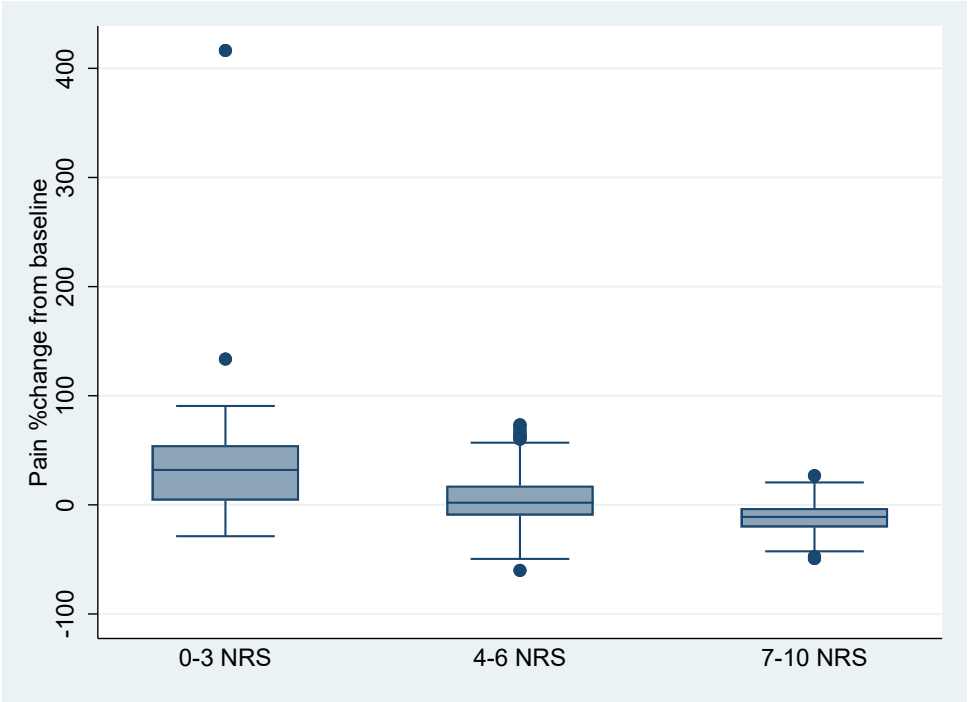
- It is an inefficient measure of treatment effect.

Vickers (2001) BMC Medical Research Methodology 1: 6

What about % change?

- In addition, % change often violates the assumptions of normality (of a normal distribution) required for the statistical tests to compare the means between groups.

- Also, the magnitude of % change depends on the baseline value.



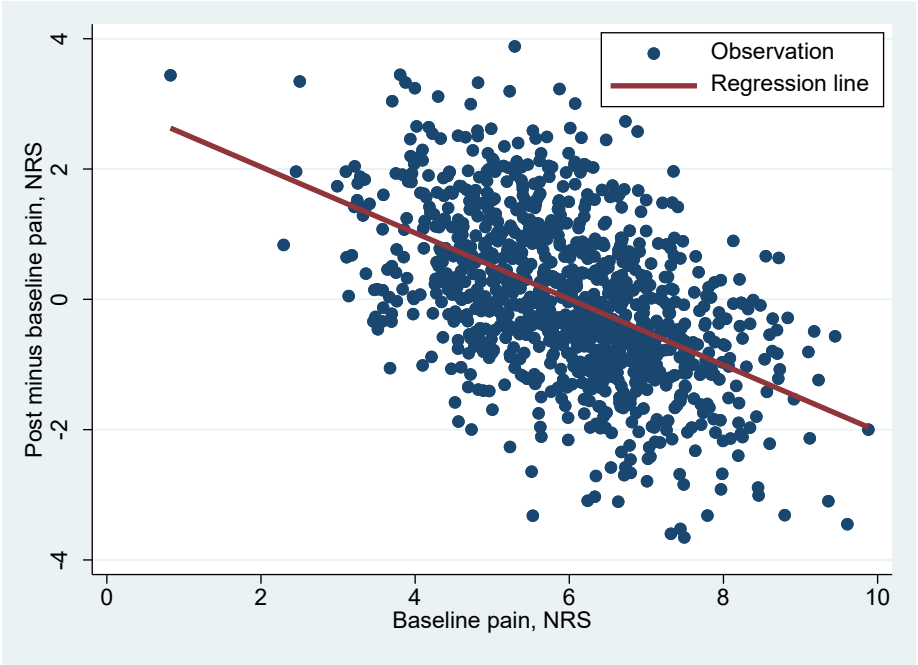
What about % change?

- What if interest lies in % change?
- If this statistic is of interest, still analyse post-intervention outcome scores with adjustment for baseline values to estimate the effect of treatment on outcome scores (95% confidence interval and p-value).
- Next, the results should be **converted** to percentage change using mean baseline and post-intervention scores for treatment and control groups.

Vickers (2001) BMC Medical Research Methodology 1: 6

Additional challenge with change scores

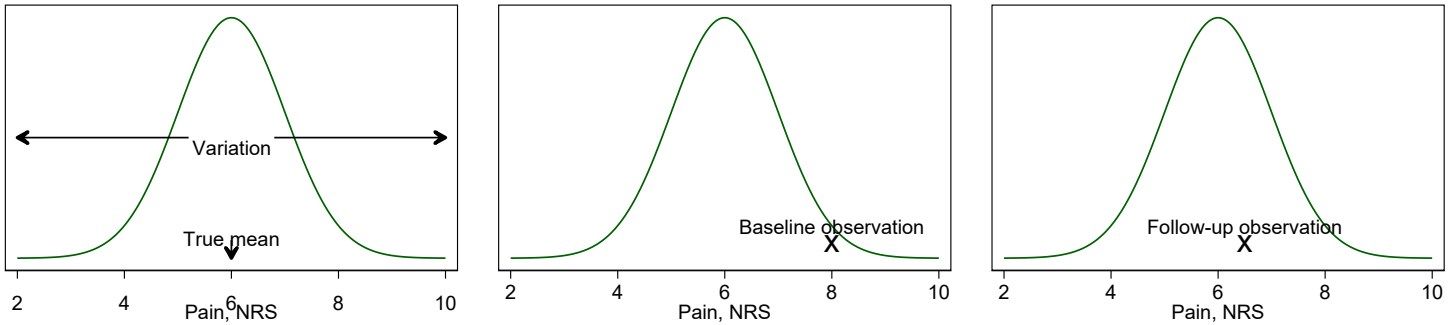
- Change scores, where there is no adjustment for baseline, do not appropriately account for regression to the mean.
- Baseline values are negatively correlated with change: patients with high pain at baseline generally improve more than those with less baseline pain.
- Points with largest change values have highest or lowest baseline.



Vickers & Altman (2001) BMJ 323: 1123

Regression to the mean

- Regression to the mean occurs when repeated measurements are made on the same subject.
- It occurs because measurements are taken with random error ("a non-systematic variation in the observed values around a true mean").
  - Repeated measures attenuate towards the average.
- Data are rarely observed without random error.



Graphical example of true mean and variation, and of regression to the mean using a Normal distribution. The distribution represents pain intensity in a single subject with a true mean of 6 NRS units and standard deviation of 2 NRS units.

Regression to the mean and single arm studies

- Consider a **single-arm study** of a new exercise treatment.
- In this study, patients are selected because their pain is higher than a certain threshold.
- As individual pain levels vary randomly over time, this could lead to the selection of patients when their pain level is above their individual long-term average.
- Subsequent reductions may be due to regression to the mean, rather than a true treatment effect.
- Without a control arm, the treatment effect cannot be separated from regression to the mean!
  - Cannot make inferences based on within-group changes.

Howard (2016) Circ Cadiovasc Qual Outcomes 9: 14-22

Extensions to multiple follow-up time points

- Some trials follow patients at multiple time-points after intervention.
- How should these be analysed?
- Baseline measurements can be considered in (constrained) longitudinal analyses.
- This is where every measurement of the outcome of interest (e.g., pain score) at all time points (e.g., baseline, follow-up 1, follow-up 2) are all considered as the outcome variable in a linear mixed model.
- In constrained longitudinal analyses, the baseline measurements are treated as equal.

Coffman (2016) BMJ Open 6: e013096

Conclusions

- ‘Adjusting for baseline’, using either the outcome at follow-up or change scores, is the best method to account for baseline values in trials with 1 baseline and 1 follow-up outcome measurement.
- Even if randomisation is designed to ensure baseline balance, adjusting for baseline protects against chance imbalance in the outcomes at baseline.
- Although change scores, without adjusting for baseline, are still commonly used, these results do not protect against regression to the mean.
- Single arm studies should be avoided as the treatment effect cannot be separated from regression to the mean.

THE UNIVERSITY OF MELBOURNE

Faculty of Medicine, Dentistry and Health Sciences

Thank you

- Website:- https://clinicalresearch.mdhs.unimelb.edu.au/
- Email:- misch-info@unimelb.edu.au

@MischHub