Statistics and Study Design

Jim Richards

Professor of Biomechanics

Allied Health Research Unit
University of Central Lancashire, Preston, UK



Aim of the Session

- Statistics is vital for the correct interpretation of research findings.
- Recognise the need for clear aims and objectives
- Identify how these can and should map to clinically important questions
- Recognise different parametric and non-parametric stats tests
- To relate study design to appropriate statistics tests

Aims and objectives

Objectives	Clinically Important Questions	Hypothesis
Objective 1	Primary clinical question	Primary
Objective 2	Secondary clinical question	Secondary
Objective 3	....	.....
Objective 4		

Spend 5 mins thinking about “possible objectives” for a new or previously published study, how these map to “clinically important questions” and “hypotheses”.

Why it is important to understand your question!!!!!



Statistical Testing of Objectives

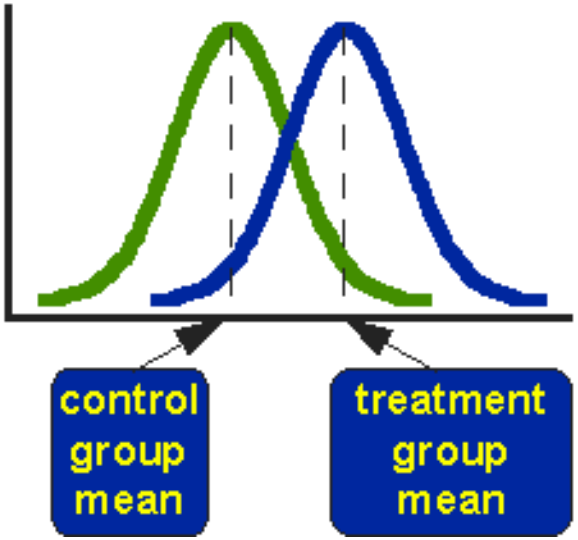


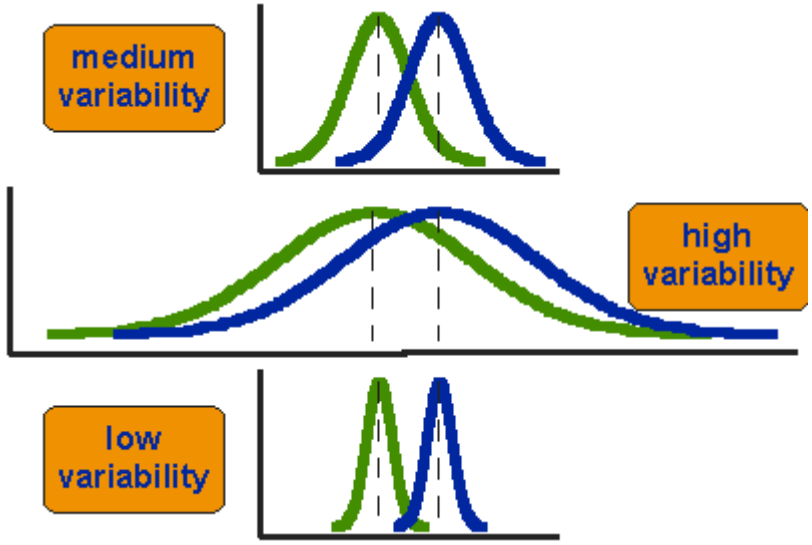
Non-parametric and Parametric

- **Non-parametric methods** consider the difference in the median and are not reliant on the data being normally distributed and are often called *distribution free*.

- **Parametric methods** consider the difference in the mean in relation to the standard deviation these assume that the distributions of the variables are known to be normally distributed.

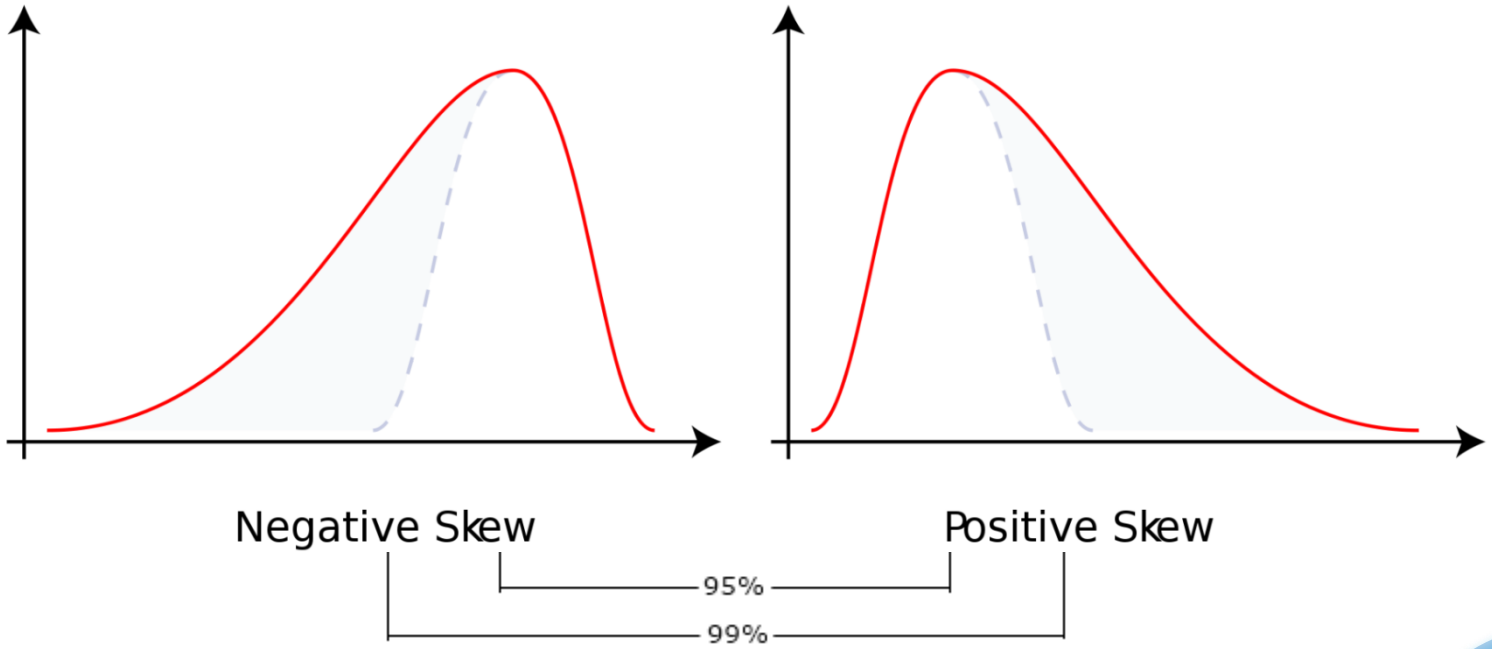
Variability and Means





Tests for normal distribution

- **Kolmogorov–Smirnov and Shapiro Wilk tests tells us if the data is normally distributed.**



If it is normally distributed we use a Parametric test
If it is not normally distributed use a non-Parametric test

Methodology and methods

- Comparison of two groups
- Comparison of more than two groups
- Type of Trial or Experiment



Comparing two data sets

Independent

- Unpaired T-tests
- Mann-Whitney

Dependant

- Paired T- tests
- Wilcoxon



Comparing two Independent groups

Unpaired T-test

- The t-test assesses whether the means of two groups are *statistically* different from each other.
- This could be used to compare the means and standard deviations of a treatment and a non-treatment group
- This analysis is appropriate whenever you want to compare the means of two groups only.

Mann-Whitney *U* test

- The Mann-Whitney ***U*** test is a non-parametric test to assesses whether the median of two groups are statistically different from each other
- It achieves this but not looking at the raw data but at the ranked data.

Raw and Ranked Data

Raw Data

patient	before	after
1	38	32
2	10	16
3	84	57
4	36	28
5	50	55
6	35	12
7	73	61
8	48	29
Median	**43**	**30.5**

Ranked Data

patient	before	after
1	9	6
2	1	3
3	16	13
4	8	4
5	11	12
6	7	2
7	15	14
8	10	5
Median	**9.5**	**9**

Comparing two Dependent groups

Paired T-test

- The t-test assesses whether the means of two assessments on the same individuals are *statistically* different from each other.
- This could be before and after a treatment.

Wilcoxon test

- The Wilcoxon signed-rank test allows a comparison of data that can be paired in the same way as in the paired T-test.
- As with the Mann-Whitney test it achieves this by ranking the data.

Do paired and unpaired show the same result?

patient	before treatment	after treatment
1	38	32
2	10	16
3	84	57
4	36	28
5	50	55
6	35	12
7	73	61
8	48	29
mean	**46.75**	**36.25**
Standard deviation	**23.2**	**19.0**

Do paired and unpaired show the same result?

Not Paired

- Unpaired t-test p = 0.339
- Mann-Whitney p = 0.3823

Paired

- Paired t-test = 0.045
- Wilcoxon p = 0.039

Comparing more than two groups

Independent

- One way ANOVA
- Kruskal-Wallis

Dependent

- Repeated Measures ANOVA
- Friedman



Comparing more than two Independent groups

One way ANOVA

- ANOVA is a general technique that can be used to test the if the means of 3 or more groups are different
- The assumption is that the sampled populations are normally distributed

Kruskal-Wallis test

- The Kruskal-Wallis test determines if the median of 3 or more groups are different
- Again this uses the same idea of ranking the scores only now 3 groups are compared
- “distribution free”.

Comparing more than two Dependent groups

Repeated Measures ANOVA

- The repeated measures ANOVA tests if the means of 3 or more observations are different
- i.e. If the same participants are tested under 3 or more conditions.

Friedman Test

- The Friedman Test is for 3 or More Dependent Samples
- Similar to the parametric repeated measures ANOVA, it is used to detect differences in treatments across multiple test attempts.
- Again this uses ranked scores.

Repeated Measures versus Not

- The follow data shows
  - testing on 8 patients before treatment
  - testing on the same 8 patients immediately after treatment
  - testing on the same 8 patients after 6 months

- This data has been analysed with repeated measures ANOVA and a one way ANOVA.

Repeated Measures versus Not

patient	before treatment	immediately after treatment	6 months after treatment
1	38	32	28
2	10	16	24
3	84	57	35
4	36	28	24
5	50	55	25
6	35	12	8
7	73	61	44
8	48	29	16

Repeated Measures versus Not

Not paired or One way ANOVA

- Kruskal-Wallis Test p = 0.0762
- One way ANOVA p = 0.094

Paired or Repeated Measures

- Friedman p = 0.0208
- Repeated Measures ANOVA p = 0.004

Post hoc testing

- One way ANOVA and repeated measures ANOVA tell us whether there is a difference between three groups or test conditions
- It does not tell us where those differences are.
- To find out which groups/test conditions are different we need to run a post hoc test

Post hoc testing

- Post hoc tests are not unlike multiple t-tests to which look at which comparisons show a difference

- Post hoc test however do this in one single step and can also take into consideration the effect of having multiple comparisons.



Post hoc testing

- As with T-tests it is very important to use the correct ANOVA, repeated measures or one way ANOVA
- The use of the wrong method can give a false result from the data
- Or may not pick up a change in the different test groups or conditions

Repeated Measures

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	1806.333	2	903.167	8.394	.004
	Greenhouse-Geisser	1806.333	1.438	1255.768	8.394	.011
	Huynh-Feldt	1806.333	1.709	1056.653	8.394	.007
	Lower-bound	1806.333	1.000	1806.333	8.394	.023
Error(factor1)	Sphericity Assumed	1506.333	14	107.595		
	Greenhouse-Geisser	1506.333	10.069	149.601		
	Huynh-Feldt	1506.333	11.966	125.880		
	Lower-bound	1506.333	7.000	215.190		

Measure:MEASURE_1

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	10.500*	4.314	.045	.300	20.700
	3	21.250*	6.611	.015	5.617	36.883
2	1	-10.500*	4.314	.045	-20.700	-.300
	3	10.750*	4.288	.041	.611	20.889
3	1	-21.250*	6.611	.015	-36.883	-5.617
	2	-10.750*	4.288	.041	-20.889	-.611

Based on estimated marginal means

*. The mean difference is significant at the .05 level.
a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

One way ANOVA

ANOVA

VAR00003

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1122.333	2	561.167	1.299	.294
Within Groups	9069.500	21	431.881		
Total	10191.833	23			

Multiple Comparisons

VAR00003
LSD

(I) pre1post2post3	(J) pre1post2post3	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	5.75000	10.39087	.586	-15.8590	27.3590
	3.00	16.50000	10.39087	.127	-5.1090	38.1090
2.00	1.00	-5.75000	10.39087	.586	-27.3590	15.8590
	3.00	10.75000	10.39087	.313	-10.8590	32.3590
3.00	1.00	-16.50000	10.39087	.127	-38.1090	5.1090
	2.00	-10.75000	10.39087	.313	-32.3590	10.8590

Trial Type

- Cross Over Trials
- Randomised Controlled Trials



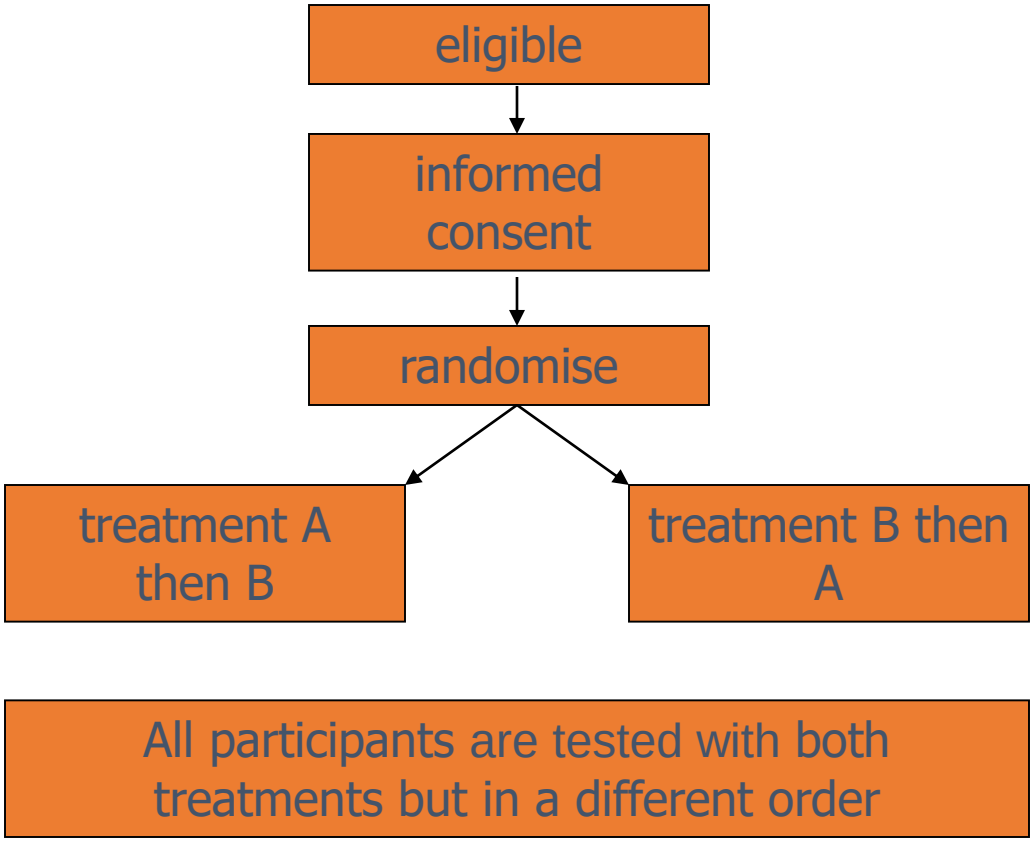
Trial Types to test hypotheses

- Efficacy and Effectiveness
- Effect of treatment in routine clinical practice
- Extent to which new treatment is equivalent to existing treatment
- Different treatments, health care practices and intervention strategies

Trial Type

- But which stats test for which occasion?
- The wrong choice of stats test can completely invalidate findings
- But how should the research student deal with this problem?

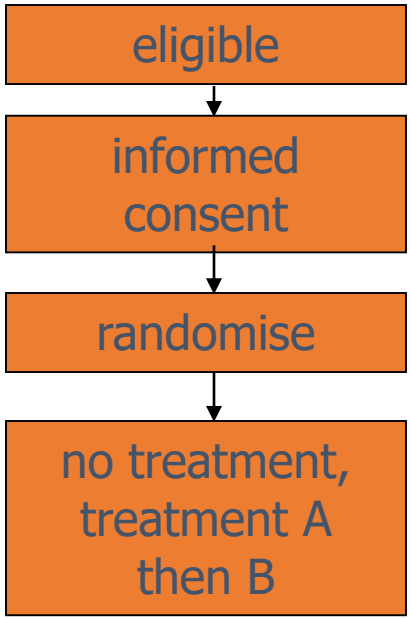
Design: Cross Over



Design: Cross Over

- What research questions can be answered with this design?
- What comparisons could you make?
- What statistics test should you use?
- What would be a better design?

Design: Cross Over with Control

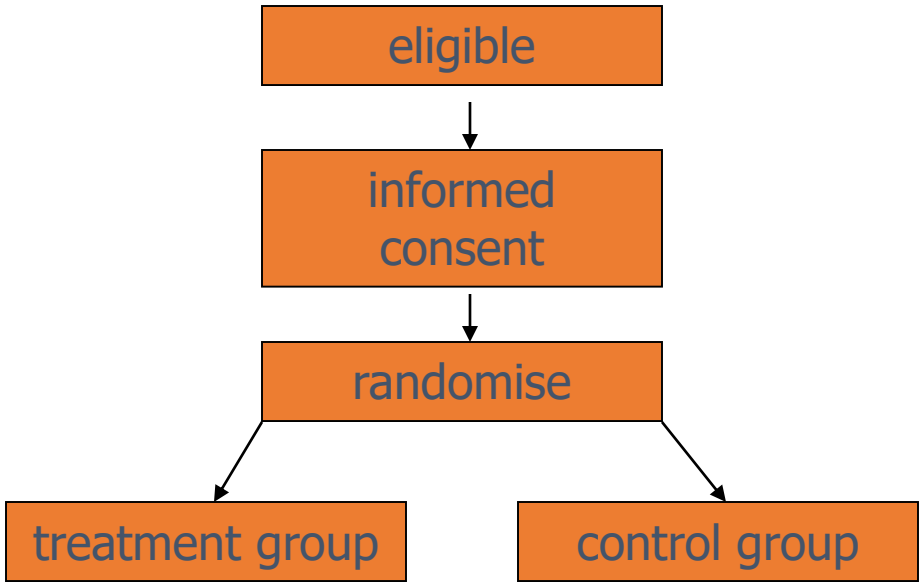


All participants are tested under all three conditions but in a different order

Design: Cross Over with Control

- What research questions can be answered with this design?
- What comparisons could you make?
- What statistics test should you use?
- What would be a better design?

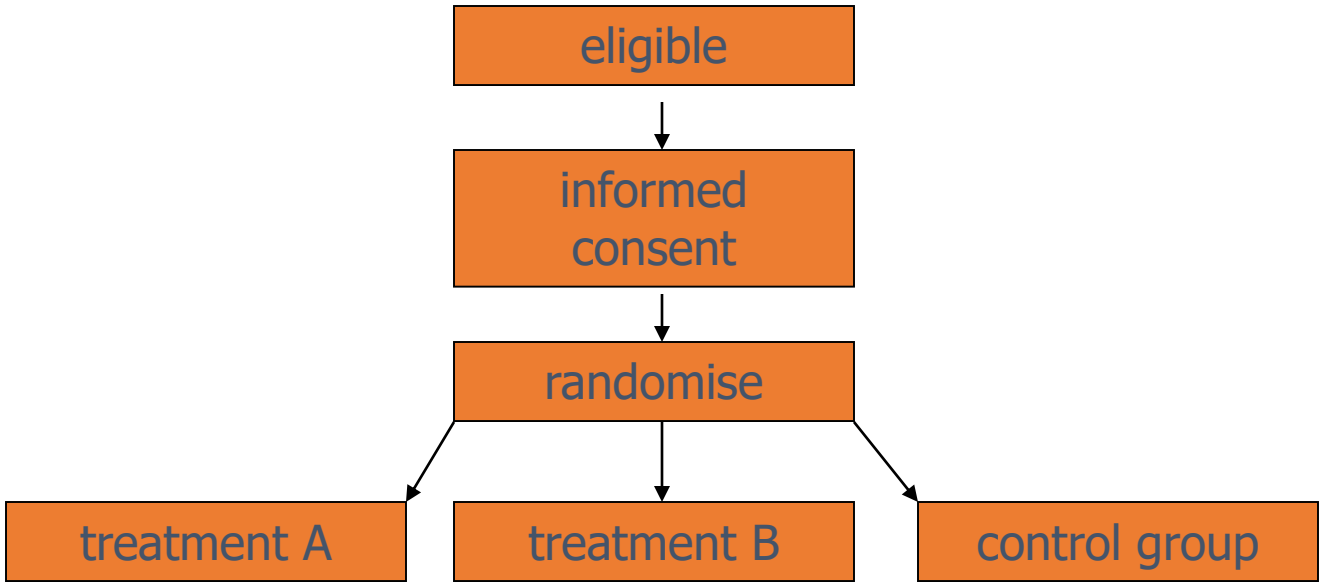
Design: Simple RCT



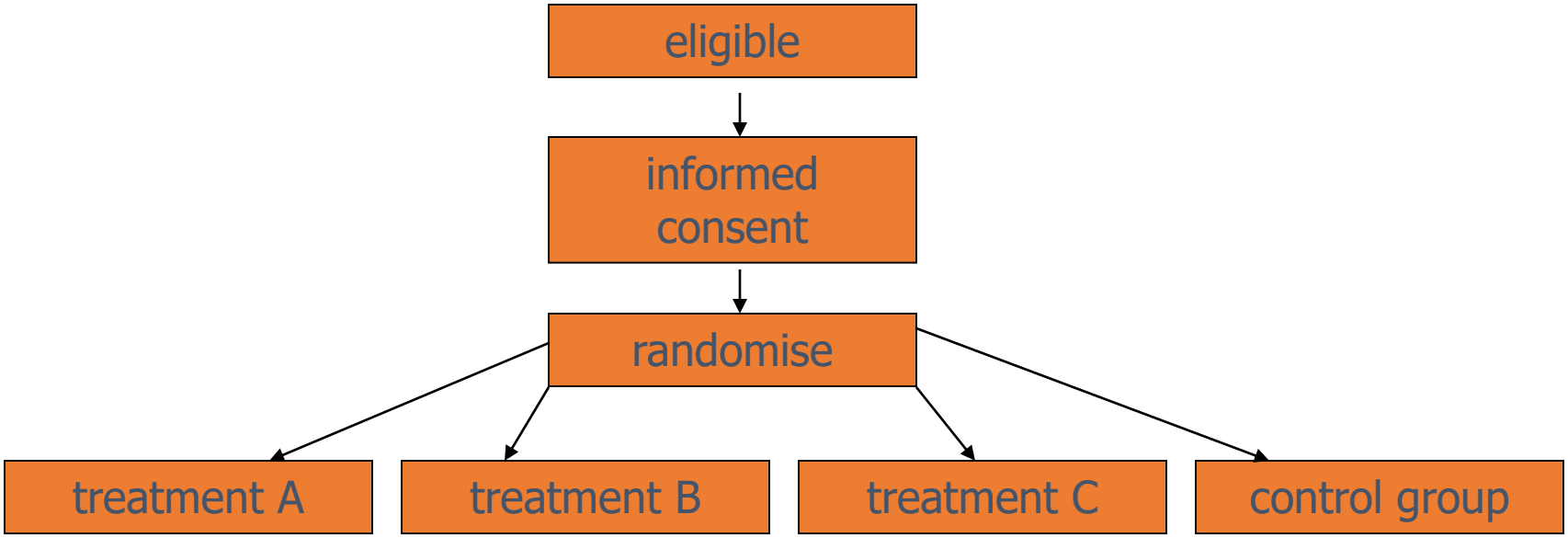
Design: Simple RCT

- What research questions can be answered with this design?
- What comparisons could you make?
- What statistics test should you use?
- What would be a better design?

Design: Three Arm RCT

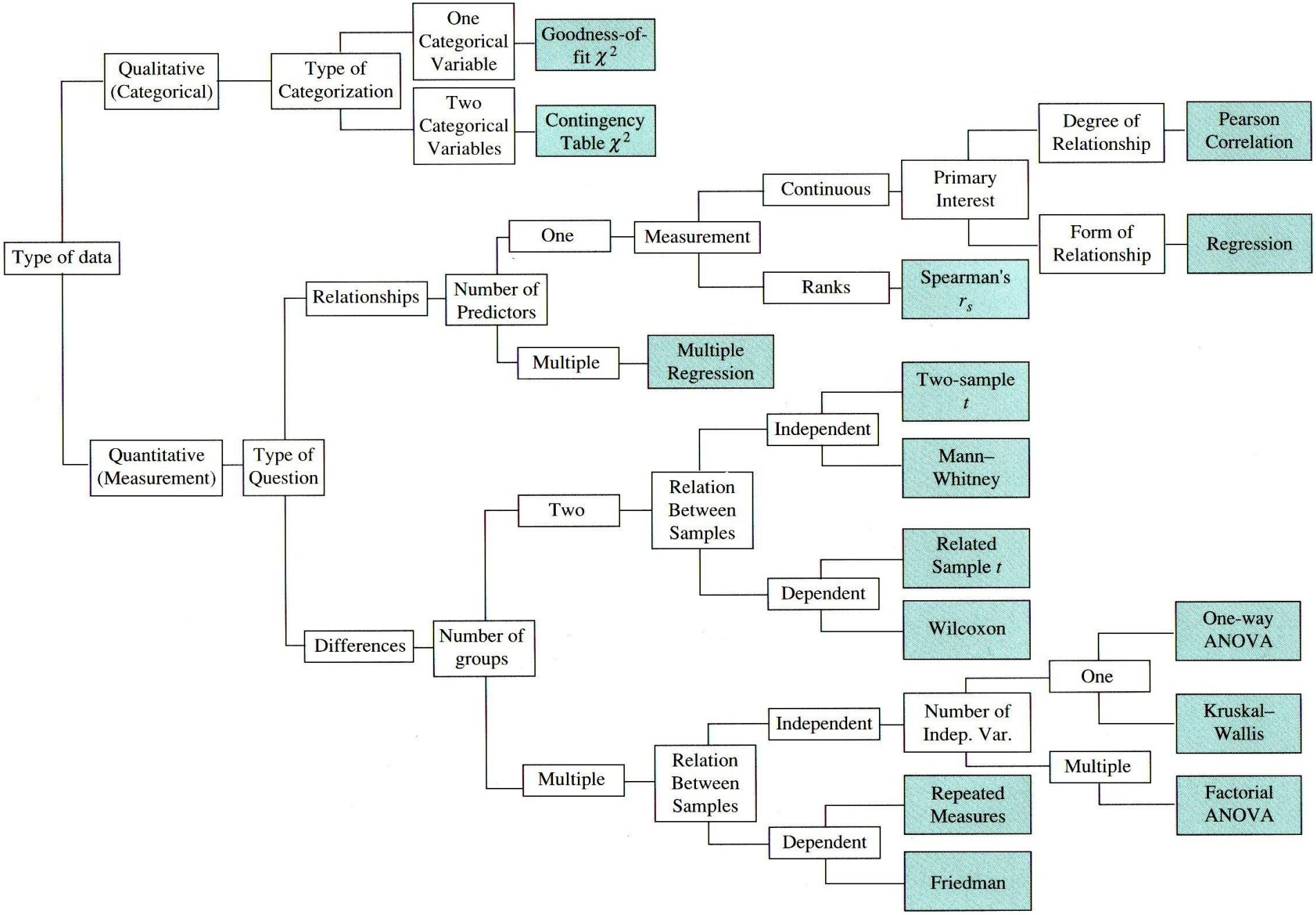


Design: Multiple Arm RCT



Design: Multiple Arm RCT

- What research questions can be answered with this design?
- What comparisons could you make?
- What statistics test should you use?


Type of data
Qualitative (Categorical)
Type of Categorization
One Categorical Variable
Goodness-of-fit χ^2
Two Categorical Variables
Contingency Table χ^2
Quantitative (Measurement)
Type of Question
Relationships
Number of Predictors
One
Measurement
Continuous
Primary Interest
Degree of Relationship
Pearson Correlation
Form of Relationship
Regression
Ranks
Spearman's r_s
Multiple
Multiple Regression
Differences
Number of groups
Two
Relation Between Samples
Independent
Two-sample t
Mann–Whitney
Dependent
Related Sample t
Wilcoxon
Multiple
Relation Between Samples
Independent
Number of Indep. Var.
One
One-way ANOVA
Kruskal–Wallis
Multiple
Factorial ANOVA
Dependent
Repeated Measures
Friedman

Examples of correct and incorrect stats use of comparison of means

- Richards et al. (2003).
- Nester et al. (2003).
- Richards et al. (2008).

Richards et al. (2003)

A comparison of knee kinematic characteristics of stroke patients and age-matched healthy volunteers

ducted with the subject waiting for a simultaneous light and auditory signal which indicated when to start the task from the standardized starting position. No other restrictions in the manner in which the subjects performed the tasks were imposed. Three trials of each movement task were performed.

The movement tasks used were adapted from Rivermead Motor Assessment.[13] These were:

1) start in standing, walk straight forward for 10 metres with no stand-by help, caliper, splint or walking aid. Subjects were asked to walk at the speed which they found most comfortable (gross function section, item 8);

Statistical analysis

The mean of the three trials for each subject was used for data analysis. The kinematic characteristics for the stroke patients and the dominant limb of healthy volunteers were compared using 2-tailed independent samples *t*-tests. For the walk and sit-to-stand tasks, the nonparetic and paretic sides were compared using two-tailed paired *t*-tests. For the step on block task the paretic lower limb of stroke patients was compared with the dominant lower limb of healthy volunteers using two-tailed independent samples *t*-tests. All calculations were conducted using SPSS version 6.1.

Nester et al. (2003)

Effect of foot orthoses on the kinematics and kinetics of normal walking gait

talar and mid tarsal joints. Rearfoot complex pronation is synonymous with foot pronation and associated with internal rotation of the leg relative to the foot. Rearfoot complex supination is synonymous with foot supination and external rotation of the leg relative to the foot.

2. Method

Fifteen subjects (8 male, 7 female) aged between 19 and 41 years and who had no musculoskeletal disorders, were recruited. Each subject performed ten gait cycles in each of three experimental conditions; shod, shod with medially wedged foot orthoses, and shod with laterally wedged foot orthoses. In each experimental condition, subjects walked at a controlled cadence of 108 steps/min and wore their own footwear.

The medially wedged orthoses comprised a 3 mm base material with a 10° wedge manufactured from high density ethylene vinyl acetate (EVA) applied under the medial aspect of the heel, continuing distally and ending prior to the first metatarsal head. The wedge was placed under the 3 mm material. In addition to the wedge, and since it is typical clinical practice, an arch filler was attached to the orthosis. This was manufactured from

dynamics. Segment masses, moments of inertia and centres of gravity were calculated using the tables of De Leva [30]. Joint moments are presented as the external moments, i.e. the moment acting on the joint. The joint moments were related to the axes of the global co-ordinate system and were normalised to each subjects body mass (kg).

The kinematic and kinetic data from each of the ten recorded gait cycles for each subject were averaged for the left and right limb separately. The reference angles (joint angles when the subject was in their relaxed standing position) were subtracted from the average data so that 0° in the angular data indicated the relaxed standing position of that subject. Data for all subjects were averaged, to produce a single data set representing the entire sample of 15 subjects. In addition, variables were derived from the data of each limb of each subject and these are listed in Table 1. The distribution of the data values about the mean value of each kinematic and kinetic variable was assessed and the data found to be suitable for parametric statistical testing. One-way ANOVA tests were performed to compare the three experimental conditions and the Tukey's Post-Hoc test was used where significant differences were found ($P < 0.05$).

Richards et al. (2008)

A Biomechanical Investigation of a Single-Limb Squat: Implications for Lower Extremity Rehabilitation Exercise

Data Processing

The raw force and movement data were exported to Visual3D (C-Motion, Inc, Germantown, MD) for processing. The movement and force data were filtered using a fourth-order, low-pass Butterworth filter with cut-off frequencies of 6 Hz and 25 Hz, respectively. The EMG data were zeroed to remove any offset and band-pass filtered with a high-pass filter of 20 Hz and a low-pass filter of 500 Hz. For the calculation of angle-specific EMG activity, data were full-wave rectified and then enveloped using a fourth-order, low-pass Butterworth filter with a

normalized based on this.

Statistical Analysis

Repeated-measures analysis of variance and post hoc pairwise comparisons were used to identify significant differences for the EMG magnitudes, iEMG values, the knee and ankle moments at 60° of knee flexion, and maximum angle of knee flexion. The Bonferroni adjustment was used to account for multiple comparisons and to reduce the possibility of type I errors. Adjusted P values were reported for comparisons of the different decline positions of the foot during the squats. The α level was set at .05.

uclan University of Central Lancashire

Allied Health Research Unit uclan University of Central Lancashire

Mixed methods

- This is also sometimes called a 2 way ANOVA with repeated measures.

- This can be used when you want to compare 2 or more groups for 2 or more time points.



Mixed methods

- The data are coded by
  - Participant number code
  - Group e.g. intervention
  - Time points

Participant number	Time point	intervention
1	1	1
2	1	2
1	2	1
2	2	2
1	3	1
2	3	2

Mixed methods

- By coding against participant number this also a dependent analysis or “repeated measures” to be conducted over the time points.

- By coding against intervention an independent analysis can be also conducted at the same time.

- This has the advantage of increasing the number of observations and therefore the power of the statistics

- However it also restricts the comparison