Summary statistics for continuous data

Tabriz University of Medical Sciences
Standard Workshop on Systematic Reviews _ October 2012

Dr. Shayesteh Jahanfar,
University of British Columbia

Outline

- identify continuous outcomes
- understand how to summarise continuous data and pool studies with:
  - measures on the same scale
  - measures on different scales
- recognise some of the challenges of continuous data

Types of data

- Binary data
- Counts of infrequent events (e.g. number of strokes)
- Short ordinal scales (e.g. pain grades: none/mild/moderate/severe)
- Long ordinal scales (e.g. disability scales)
- Continuous data (e.g. blood pressure)
- Censored data (e.g. survival times)

What are continuous data?

- data with an infinite number of values that are equally spaced
- example: height - it can be measured along a numerical continuum of centimetres, metres or inches, feet
  - a person can be 175.24678cm tall, assuming the measurement instrument is accurate enough
  - the difference between 160 and 161cm, and 180 and 181cm, is the same

Long ordinal scales

- sometimes treated as continuous data
- but not true continuous because
  - they have a finite number of distinct values
  - there are gaps in the continuum
- have multiple, ordered categories which imply magnitude
  - e.g. one category is greater or lesser than another
- spacing between categories is not numerically equivalent
- approach ‘continuous’ with increasing categories

What continuous data can we combine?

- data represent continuous measures
- the mean value is in the middle (distribution is roughly symmetrical)
- measurements are made on all participants (not censored or survival type data)
- data are available for both groups in each trial

What data is needed?

	Mean	SD	Sample size
Treatment	m_t	sd_t	n_t
Control	m_c	sd_c	n_c

Meta-analysis of continuous data

- calculate a single summary statistic to represent the effect found in each study
- Summary statistics combined in meta-analysis
- 2 options
  - mean difference
  - standardised mean difference

Mean difference

- outcomes measured in same unit using same scale (e.g. blood pressure as mmHg)
- pooled analysis in “natural units” and therefore easy to interpret
- studies weighted according to the inverse of the variance (a function of size and SD)

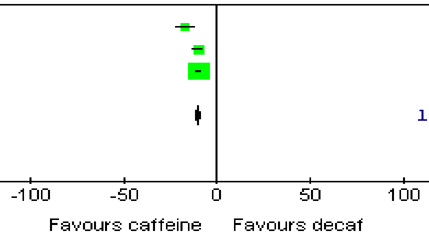
MD = mean on treatment – mean on control

Mean difference: example

Review: Caffeine for daytime 'sluggishness'. (version with data)
Comparison: 01 Caffeinated Coffee versus Decafeinated Coffee
Outcome: 03 Irritability at 30 minutes - INAS scale (1-50, high score worse)

Study or sub-category	Caffeine N	Caffeine Mean (SD)	Decaf N	Decaf Mean (SD)	WMD (fixed) 95% CI	Weight %	WMD (fixed) 95% CI
Nescafe 1998	68	19.00(15.50)	64	36.00(17.30)		4.00	-17.00 [-22.62, -11.38]
Harris Hudsons 2002	65	20.00(9.10)	67	30.00(8.60)		13.82	-10.00 [-13.02, -6.98]
Andronicus 2004	40	20.00(2.40)	40	30.00(3.20)		82.17	-10.00 [-11.24, -8.76]
Total (95% CI)	173		171			100.00	-10.28 [-11.40, -9.16]

Test for heterogeneity: Chi² = 5.73, df = 2 (P = 0.06), I² = 65.1%
Test for overall effect: Z = 17.93 (P < 0.00001)



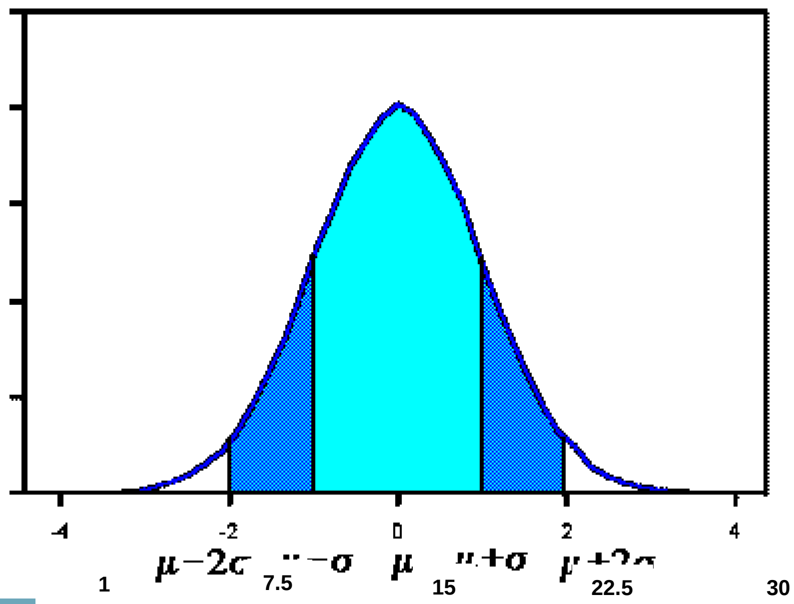
Standardised mean difference

- Outcome is same concept measured on different scales, the values must be transformed to a common scale before pooling
- Sometimes scale factors are known and transformations are made directly (e.g weight)
- Standardised mean difference calculated as:

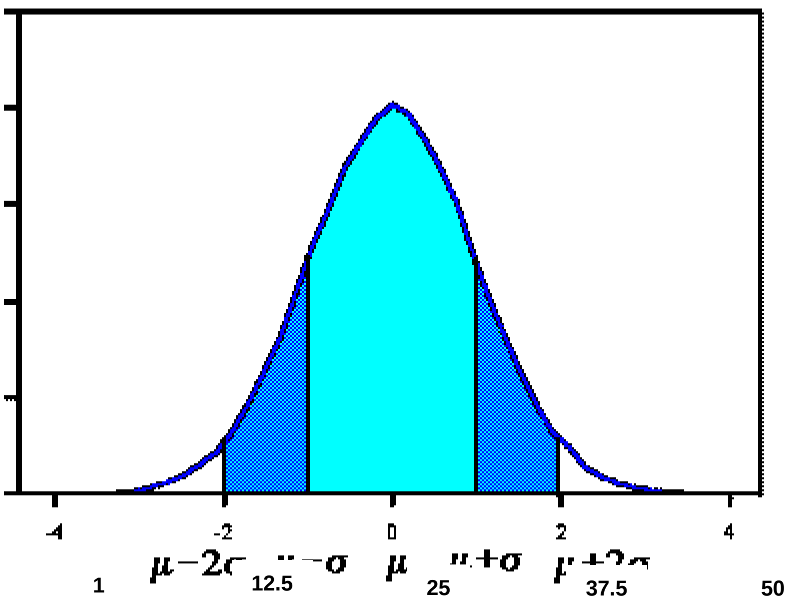
$$\frac{\text{Difference in means between groups}}{\text{Average standard deviation}}$$

Standardised mean difference

Beck Irritability Scale (1-30)



Irritability Negativity Affectivity Subscale (1-50)



Different scales but averages mean the same thing
(i.e. average person is just as irritable!)

Measurements on different scales

Comparing irritability at 30 minutes between caffeinated coffee and decafe coffee

Trial	Caffeinated N. mean (SD)	Decafe N. mean (SD)	Irritability scale
Moccona 1998	15 23.0 (15.1)	17 31.0 (15.2)	INAS
Nescafe 1998	68 19.0 (15.5)	64 36.0 (17.3)	INAS
Piazza D'oro 2003	35 21.0 (3.2)	37 10.0 (4.20)	BII

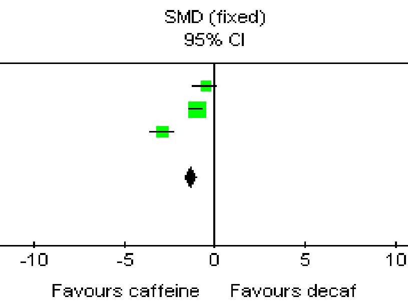
High scores on the Beck Irritability Scale (BII) (1-30) good outcomes, while high scores on the Irritability Negative Affectivity Subscale (INAS) (1-50) are poor outcomes

SMD: example

Review: Caffeine for daytime 'sluggishness'. (version with data)
Comparison: 01 Caffeinated Coffee versus Deca
Outcome: 06 Irritability at 30 minutes

Study or sub-category	Caffeine N	Caffeine Mean (SD)	Decaf N	Decaf Mean (SD)	Weight %	SMD (fixed) 95% CI
Moccona 1998	15	23.00(15.10)	17	31.00(15.20)	16.99	-0.51 [-1.22, 0.19]
Nescafe 1998	68	19.00(15.50)	64	36.00(17.30)	64.18	-1.03 [-1.39, -0.67]
Piazza D'Oro 2003	35	-21.00(3.20)	37	-10.00(4.20)	18.83	-2.90 [-3.58, -2.23]
Total (95% CI)	118		118		100.00	-1.30 [-1.59, -1.00]

Test for heterogeneity: Chi² = 28.72, df = 2 (P < 0.00001), I² = 93.0%
Test for overall effect: Z = 8.71 (P < 0.00001)



RevMan exercise

Change vs endpoint scores

Start of study

End of study

atment group

Score T0

Change in score T

Score T1

Difference in mean change scores

Difference in mean end point scores

ontrol group

Score C0

Change in score C

Score C1

Problems with MD and SMD

- what constitutes a clinically important change?
- restrictive eligibility criteria results in smaller standard deviations; therefore these trials given more weight
- mean difference
  - measurements on the same scale are not always comparable (e.g. health care costs in different places, process of care measures)
- standardised mean difference
  - difficult to interpret outcomes in units of SD, but can transform back to units of the scale
  - estimates of variation may not always be comparable making the SD a poor scaling factor

Take home message

- pooling continuous data – use mean difference or standardised mean difference
- check data for skewness
- can calculate SDs from other statistics
- can use either endpoint or change scores