Meta-analysis of continuous data: Final values, change scores, and ANCOVA

Jo McKenzie

School of Public Health and Preventive Medicine, Monash University, Australia

joanne.mckenzie@med.monash.edu.au

Cochrane Statistical Methods Training Course: Addressing advanced issues in meta-analytical techniques

4-5th March 2010, Cardiff, UK

MONASH University

Outline of the presentation

- Common questions about final values (FV), change scores (CS), and ANCOVA in meta-analysis.
- Analysis of a randomised trial:
  - Example demonstrating how the analytical methods compare.
  - Properties of the estimators:
    - Conditional and unconditional on baseline imbalance,
    - Preferred estimator.
- Meta-analysis:
  - A small simulation study:
    - Key points.
  - Advice from the Cochrane Handbook for Systematic Reviews of Interventions.
  - A proposed hierarchy of options.
  - An example.
  - Some final thoughts.

Common questions: SMG list (October 2007)

"In a Cochrane Workshop I attended <*snip*> we were told to use the post test scores and SDs of continuous variables. When we asked about what to do in case there is a significant difference in pretest scores, they stated that if proper randomization is used, pretest scores should be similar and therefore we should only concern ourselves with the posttest scores. However, we all know that is not always the case in smaller studies."

"I just spoke with <*snip*> and he strongly recommends using the change scores as this method better accounts for potential pretest differences."

"… What do you recommend to authors in your CRG: use of final value or use of change scores?"

Common questions

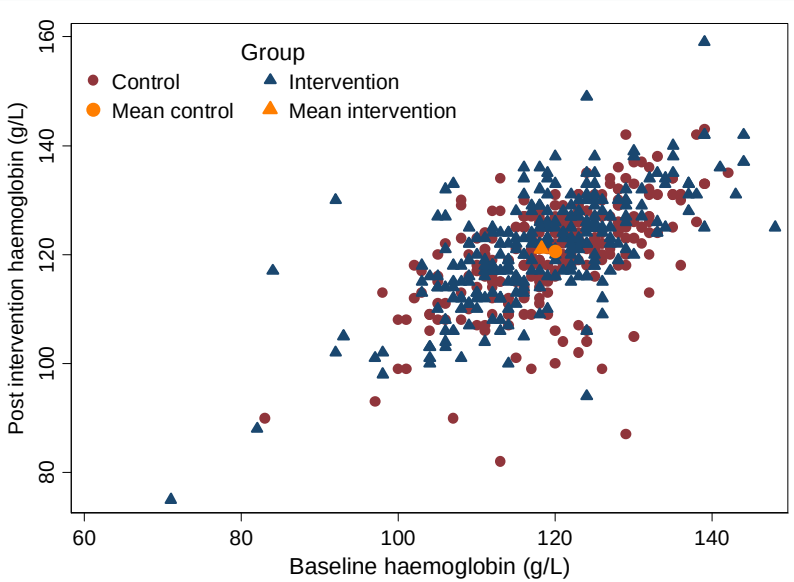
- There were 16 posts to the list following the initial post with the conclusion:

  "… no consensus exists in the SMG"

Example comparing estimators

- Randomised trial carried out in the Ubon Ratchathani province NE Thailand.
- Aimed to test the efficacy of a seasoning powder fortified with micronutrients.
- Groups:
  - Intervention: fortified seasoning powder added to instant wheat noodles or rice,
  - Control: unfortified seasoning powder added to instant wheat noodles or rice.
- Data collected at baseline and follow-up (31 weeks).
- Primary outcome was anaemia (defined from the continuous variable haemoglobin).

Post intervention haemoglobin vs baseline haemoglobin



Estimators

- Three commonly used estimators are the final value (FV), change score (CS) and analysis of covariance (ANCOVA) estimator.

$$\hat{\theta}_{FV} = \bar{y}_{\text{int}} - \bar{y}_{ctrl}$$

$$\hat{\theta}_{CS} = (\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - (\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

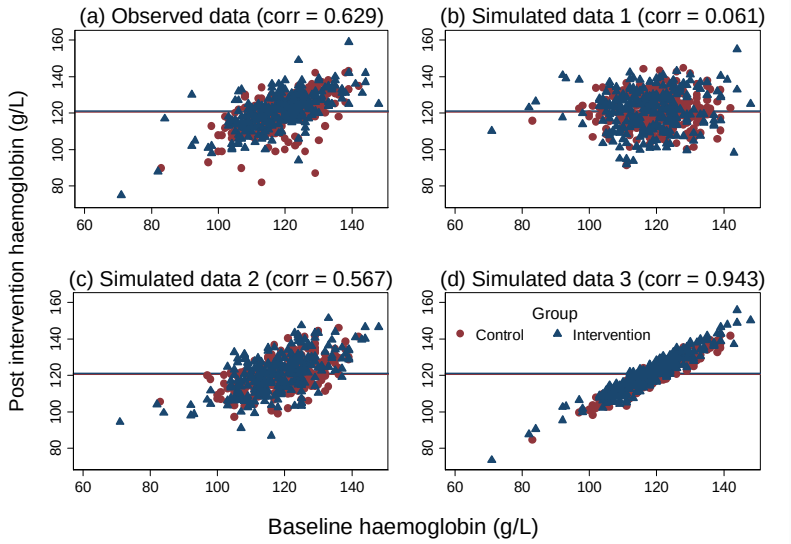
$$\hat{\theta}_{ANCOVA} = (\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - \beta(\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

where $\beta = \rho \frac{\sigma_y}{\sigma_x}$

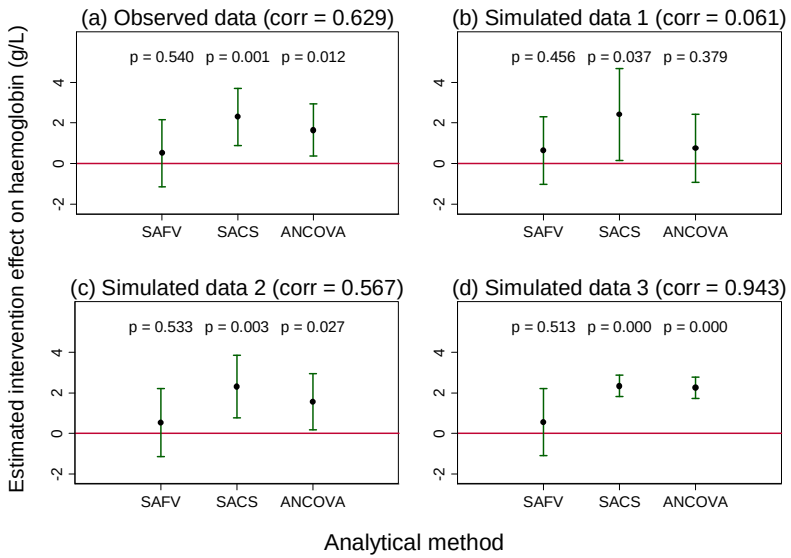
Observed and simulated data sets: summary statistics

Dataset	Observed correlation	Follow-up			
		Intervention group		Control group	
		Mean	SD	Mean	SD
Observed data	0.629	121.0	10.1	120.5	9.5
Simulated data 1	0.061	121.2	10.8	120.6	8.8
Simulated data 2	0.567	121.2	10.8	120.6	8.8
Simulated data 3	0.943	121.1	10.5	120.5	9.0

Scatter plots of post intervention haemoglobin vs baseline haemoglobin for observed and simulated data sets



Estimated intervention effects (95% CIs) calculated using different analytical methods for the four data sets



Some observations from comparing the analytical methods

- Estimates of intervention effect:
  - For a particular data set, the three estimators can produce different estimates of intervention effect.
    - When the correlation is close to 0, $\hat{\theta}_{ANCOVA} \approx \hat{\theta}_{FV}$.
    - When the correlation is close to 1, $\hat{\theta}_{ANCOVA} \approx \hat{\theta}_{CS}$.
  - Over the data sets (varying correlation), the ANCOVA estimate varies, however, a simple analysis of either change scores (SACS) or final values (SAFV) does not.
- Standard errors:
  - The standard error (SE) of the FV estimator remains the same over all data sets.
  - As the correlation increases, the SE of the CS estimator decreases. When the correlation < 0.5, the SE of CS estimator is > SE of the FV estimator. This is reversed when the correlation is > 0.5.
  - For a particular data set, the SE of the ANCOVA estimate is smaller compared to the SEs of FV and CS.

Relationship between the estimators

$$\hat{\theta}_{ANCOVA} = (\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - \beta(\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

- Assuming $\sigma_y^2 = \sigma_x^2$:
  - When the correlation is close to 0, $\beta \approx 0$; $\hat{\theta}_{ANCOVA} \approx \hat{\theta}_{FV}$
  - When the correlation is close to 1, $\beta \approx 1$; $\hat{\theta}_{ANCOVA} \approx \hat{\theta}_{CS}$
- When there is minimal baseline imbalance; the three methods produce similar estimates since $(\bar{x}_{\text{int}} - \bar{x}_{ctrl}) = 0$.

Variance of the estimators

- Assuming $\sigma^2_{y,\text{int}} = \sigma^2_{y,ctrl} = \sigma^2_{x,\text{int}} = \sigma^2_{x,ctrl} = \sigma^2$

$$\sigma^2_{FV} = \frac{2\sigma^2}{n}$$

$$\sigma^2_{CS} = \frac{4\sigma^2}{n}(1-\rho)$$

$$\sigma^2_{ANCOVA} = \frac{2\sigma^2}{n}(1-\rho^2)$$

Statistical properties of the estimators: unconditional inference

- How do the estimators perform over hypothetical repetitions of randomised trials where baseline imbalance randomly varies?
- Bias
  - SAFV, SACS, and ANCOVA are unbiased.
  - Type I error rates will be as designated.
- Precision
  - ANCOVA is (generally) more efficient compared with SAFV, or a SACS.

Statistical properties of the estimators: conditional inference

- How do the estimators perform over hypothetical repetitions of randomised trials with the same baseline imbalance?
- Bias
  - ANCOVA is conditionally unbiased.
  - SAFV is conditionally biased:

$$\mathrm{E}(\bar{y}_{\text{int}} - \bar{y}_{ctrl} \mid \bar{x}_{\text{int}} - \bar{x}_{ctrl}) = \theta + \rho\left(\frac{\sigma_y}{\sigma_x}\right)(\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

  - SACS is conditionally biased:

$$\mathrm{E}((\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - (\bar{x}_{\text{int}} - \bar{x}_{ctrl}) \mid \bar{x}_{\text{int}} - \bar{x}_{ctrl}) = \theta + \left(\rho\frac{\sigma_y}{\sigma_x} - 1\right)(\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

Statistical properties of the estimators: conditional inference

- Precision
  - ANCOVA is (generally) more efficient compared to SAFV, or a SACS.

Unconditional or conditional inference?

- Some debate over which type of statistical inference is important [Senn 1989; Overall 1992; Wei 2001].
- "If I am cruising at 10,000m above sea level in mid-Atlantic and the captain informs me that three engines are on fire, I can hardly console myself with the thought that, on average, air travel is very safe" [Senn 1997].
- Most trialists are concerned about the potential impact of any observed baseline imbalance on their estimate of intervention effect. Not reassured by the knowledge that, on average, the analytical method will provide an unbiased estimate of intervention effect.
- **ANCOVA the preferred analytical method both conditionally and unconditionally.**

What about meta-analysis? A small simulation study (1)

- Overview:
  - Number of trials per meta-analysis randomly selected from U(3, 5).
  - These were analysed using all SAFV, all SACS, all ANCOVA, or a random mix of the three analytical methods.
  - Results pooled using the standard inverse variance fixed effect model.
  - For each simulation scenario there were 50,000 replicates.

A small simulation study (2)

- Number of participants per trial randomly selected from N(20, 9).
- Baseline and follow-up scores randomly sampled from a bivariate normal distribution for both the intervention and control groups (no intervention effect).

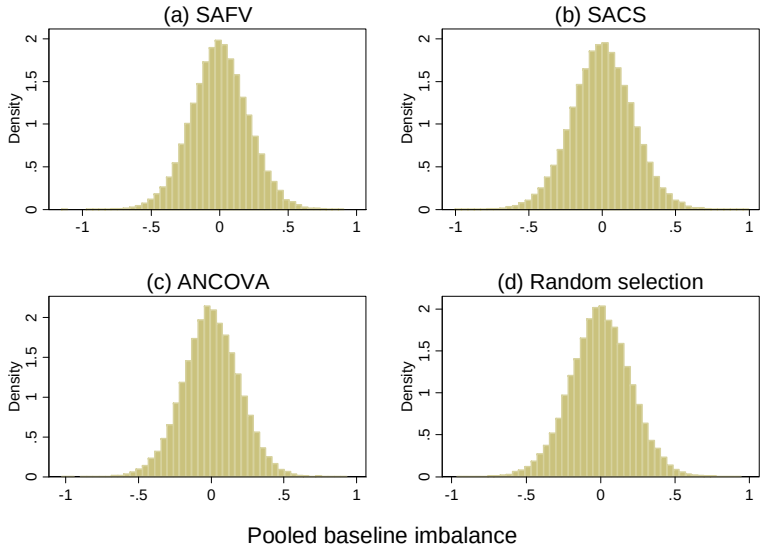
$$\begin{pmatrix} x \\ y \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

- Assuming

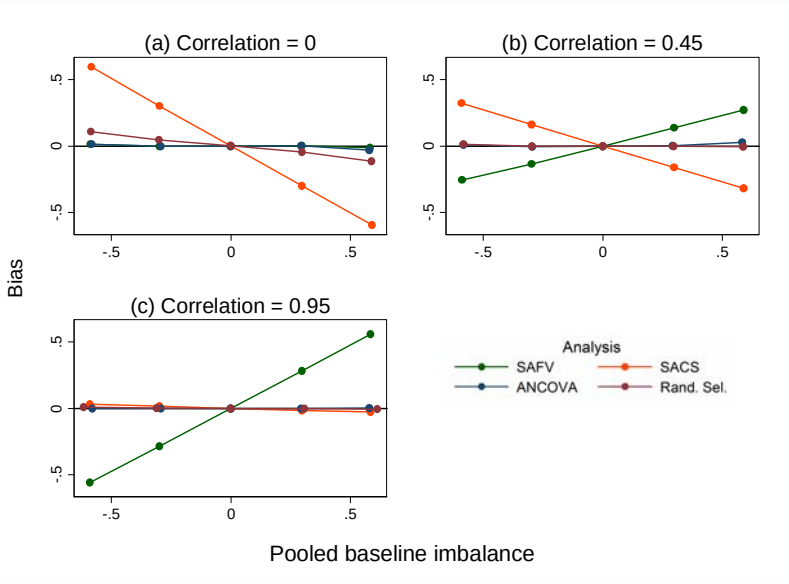
$$\sigma^2_{y,\text{int}} = \sigma^2_{y,ctrl} = \sigma^2_{x,\text{int}} = \sigma^2_{x,ctrl} = 1$$

- Seven correlations: 0, 0.05, 0.25, 0.45, 0.65, 0.85, 0.95.

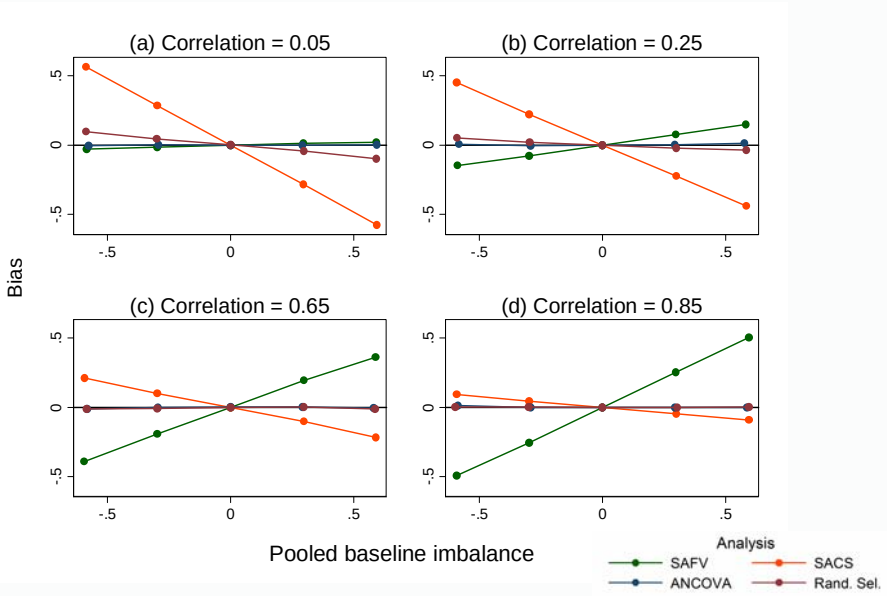
Distribution of pooled baseline imbalance (corr = 0.45)



Bias in the pooled estimate of intervention effect conditional on baseline imbalance (graph 1)



Bias in the pooled estimate of intervention effect conditional on baseline imbalance (graph 2)



A small simulation study: Key points (1)

- Pooled baseline imbalance is not often a problem (for trials with adequate allocation concealment):
  - Less likely to be a problem as the number of trials increases, or the sample size of the component trials increases, or both.
- When there is no, or minimal, baseline imbalance:
  - Pooling using all SACS, all SAFV, all ANCOVA, or a random selection of analytical methods will produce an unbiased pooled estimate of intervention effect.
  - ANCOVA more efficient than SAFV or SACS.
  - Efficiency of pooling using all SAFV compared with all SACS dependent on the 'overall' correlation between baseline and follow-up.

A small simulation study: Key points (2)

- When there is pooled baseline imbalance:
  - Pooling using all ANCOVA produces an unbiased pooled estimate of intervention effect. Guards against baseline imbalance.
  - Pooling using all SAFV will produce a biased pooled estimate of intervention effect (except if the correlation tends to be close to 0).
  - Pooling using all SACS will produce a biased pooled estimate of intervention effect (except if the correlation tends to be close to 1).
  - Pooling using a random selection of analytical methods will produce a biased pooled estimate of intervention effect.

Meta-analysis in the real world

- In publications of trials:
  - Generally only one type of analysis will be reported.
  - Less frequently, an alternative analysis will be provided, or enough information to allow an alternative analysis to be performed.
- The analytical method used may (is likely to?) vary across trials.

Advice from the Handbook (1)

- Relevant sections: 9.4.5, 16.1.3.
- "The preferred statistical approach to accounting for baseline measurements of the outcome variable is to include the baseline outcome measurements as a covariate in a regression model or analysis of covariance (ANCOVA). These analyses produce an 'adjusted' estimate of the treatment effect together with its standard error. These analyses are the least frequently encountered, but as they give the most precise and least biased estimates of treatment effects they should be included in the analysis when they are available." [Section 9.4.5.2]

Advice from the Handbook (2)

- "In practice an author is likely to discover that the studies included in a review may include a **mixture of change-from-baseline and final value scores. However, mixing of outcomes is not a problem when it comes to meta-analysis of mean differences. There is no statistical reason why studies with change-from-baseline outcomes should not be combined in a meta-analysis with studies with final measurement outcomes when using the (unstandardized) mean difference method in RevMan. In a randomized trial, mean differences based on changes from baseline can usually be assumed to be addressing exactly the same underlying intervention effects as analyses based on final measurements. That is to say, the difference in mean final values will on average be the same as the difference in mean change scores.** If the use of change scores does increase precision, the studies presenting change scores will appropriately be given higher weights in the analysis than they would have received if final values had been used, as they will have smaller standard deviations." [Section 9.4.5.2]

Concern regarding this advice

- The potential of within-study selective reporting bias has been defined as:
  - "… selection, on the basis of the results, of a subset of analyses undertaken to be included in a study publication." [Williamson 2005]
- As shown earlier in the Thailand trial and simulated data sets, the analytical methods will often produce different estimates of intervention effect.
  - Selection of the most favourable estimate is likely to result in a biased pooled estimate of intervention effect and inflated type I error rates.

Meta-analysis options: A proposed hierarchy of options

- Please see handout.

1. Obtain individual patient data for each trial and reanalyse these data.
2. Pool using only ANCOVA results. For each trial recreate the ANCOVA estimate from the summary statistics provided. This will generally involve imputing only the correlation.
3. Pool using results from only one analytical method (SACS or SAFV).
4. Pool using a mix of results from different analytical methods.

An example: Calcium supplementation on body weight

- Trowman R, Dumville JC, Hahn S, Torgerson DJ. A systematic review of the effects of calcium supplementation on body weight. Br J Nutr 2006;95(6):1033-8.
- Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. J Clin Epidemiol 2007;60(12):1229-33.

Example 2: Study characteristics (modified table 1) (Trowman, 2006)

Study	Number of participants	Age*	Sex	Intervention (Ca concentration)	Length of follow-up	Country
Chee et al. (2003)	173	58·9	Female (postmenopausal)	Ca supplement (1200 mg/d)	24 months	Malaysia
Jensen et al. (2001)	52	NA	Female (obese postmenopausal)	Ca supplement (1000 mg/d)	26 weeks	Denmark
Lau et al. (2001)	185	57·0	Female (postmenopausal)	Ca supplement (800 mg/d)	24 months	China
Reid et al. (2002)	223	72·0	Female (postmenopausal)	Ca supplement (1000 mg/d)	24 months	New Zealand
Shapses et al. (2004)	36	59·3	Female (obese postmenopausal)	Ca supplement (1000 mg/d)	25 weeks	USA
Shapses et al. (2004)	30	56·0	Female (obese postmenopausal)	Ca supplement (1000 mg/d)	25 weeks	USA
Shapses et al. (2004)	42	41·0	Female (obese postmenopausal)	Ca supplement (1000 mg/d)	25 weeks	USA
Winters-Stone & Snow (2004)	23	24·8	Female (athletes)	Ca supplement (1000 mg/d)	12 months	USA
Zemel et al. (2004)	41	46	Mixed (obese)	Calcium supplement (800 mg/d)	24 weeks	USA

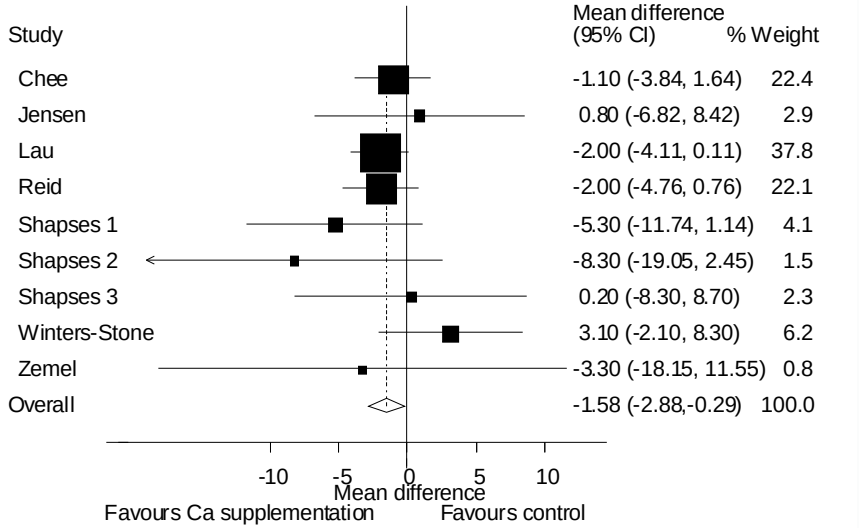
NA, not available
*** Mean age. When age was reported separately by subgroups, the mean between the groups was calculated.**

Calcium supplementation on body weight (Trowman, 2006)

Trial	Year	Baseline (weight kg)				Follow-up (weight kg)		Change (weight kg)	
		Intervention		Control		Intervention	Control	Intervention	Control
		n	Mean (SD)	n	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Chee	2003	91	56.1 (8.9)	82	57.2 (9.4)	*56.1 (?)*	*57.4 (?)*	0.0 (2.6) [a]	0.2 (2.6) [a]
Jensen	2001	25	94.6 (14.0)[a]	27	93.8 (14.0)[a]	89.0 (12.7)[a]	89.1 (14.7)[a]	*-5.6 (?)*	*-4.7 (?)*
Lau	2001	95	56.9 (7.1)	90	58.9 (7.5)	*57.4 (?)*	*58.6 (?)*	0.5 (2.6)[a]	-0.3 (2.7)[a]
Reid	2002	111	66.0 (10.0)	112	68.0 (11.0)	*65.7 (?)*	*67.9 (?)*	-0.3 (1.8)	-0.1 (2.4)
Shapses 1[c]	2004	17	84.1 (9.4)	19	89.4 (10.3)	*77.1 (?)*	*82.1 (?)*	-7.0 (4.6)	-7.3 (5.3)
Shapses 2[c]	2004	11	85.9 (9.2)	11	94.2 (15.7)	*79.2 (?)*	*86.6 (?)*	-6.7 (2.6)	-7.6 (5.7)
Shapses 3[c]	2004	18	93.7 (13.6)	24	93.5 (14.3)	*87.0 (?)*	*89.2 (?)*	-6.7 (5.5)	-4.3 (3.5)
Winters-Stone	2004	13	57.2 (4.9)	10	54.1 (7.2)	56.3 (4.3)	54.8 (7.2)	*-0.9 (?)*	*0.7 (?)*
Zemel	2004	11	99.8 (14.9)	10	103.1 (19.3)	*91.2 (?)*	*96.5 (?)*	-8.6 (5.3)[a]	-6.6 (8.2) [a]

a Calculated from the standard error
b Follow-up sample size ntrt = 24 and nctrl = 24
c Shapses *et al* (Shapses *et al*, 2004) report on three randomised controlled trials.
Trials 1, 2, and 3 include postmenopausal women, postmenopausal women special diet, and premenopausal women respectively.

Meta-analysis of difference in means at baseline



Calcium supplementation on body weight (Trowman, 2006)

Trial	Year	Baseline (weight kg)				Follow-up (weight kg)		Correlation between baseline and follow-up	
		Intervention		Control		Intervention	Control	Intervention	Control
		n	Mean (SD)	n	Mean (SD)	Mean (SD)	Mean (SD)	Correlation	Correlation
Chee	2003	91	56.1 (8.9)	82	57.2 (9.4)	*56.1* ***(8.9)***	*57.4* ***(9.4)***		
Jensen	2001	25	94.6 (14.0)[a]	27	93.8 (14.0)[a]	89.0 (12.7)[a]	89.1 (14.7)[a]		
Lau	2001	95	56.9 (7.1)	90	58.9 (7.5)	*57.4* ***(7.1)***	*58.6* ***(7.5)***		
Reid	2002	111	66.0 (10.0)	112	68.0 (11.0)	*65.7* ***(10.0)***	*67.9* ***(11.0)***		
Shapses 1[c]	2004	17	84.1 (9.4)	19	89.4 (10.3)	*77.1* ***(9.4)***	*82.1* ***(10.3)***		
Shapses 2[c]	2004	11	85.9 (9.2)	11	94.2 (15.7)	*79.2* ***(9.2)***	*86.6* ***(15.7)***		
Shapses 3[c]	2004	18	93.7 (13.6)	24	93.5 (14.3)	*87.0* ***(13.6)***	*89.2* ***(14.3)***		
Winters-Stone	2004	13	57.2 (4.9)	10	54.1 (7.2)	56.3 (4.3)	54.8 (7.2)		
Zemel	2004	11	99.8 (14.9)	10	103.1 (19.3)	*91.2* ***(14.9)***	*96.5* ***(19.3)***		

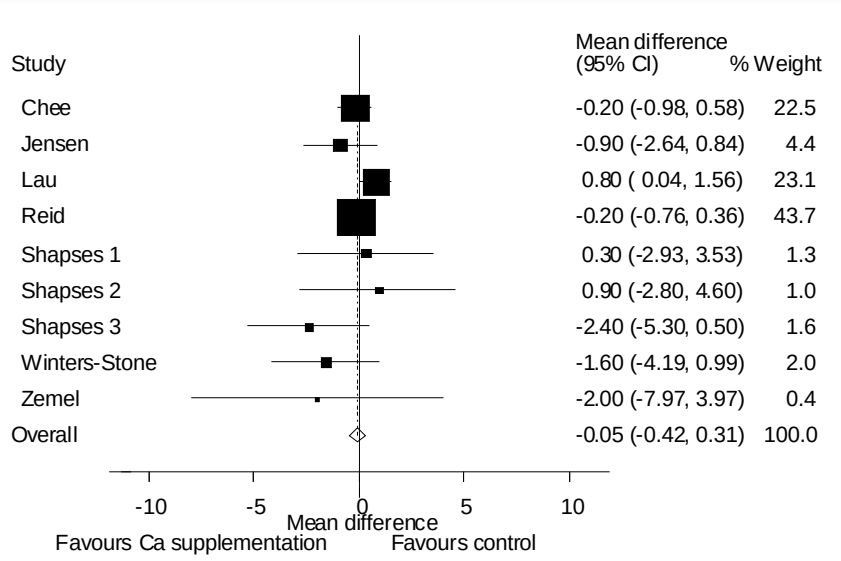
a Calculated from the standard error
b Follow-up sample size ntrt = 24 and nctrl = 24
c Shapses *et al* (Shapses *et al*, 2004) report on three randomised controlled trials.
Trials 1, 2, and 3 include postmenopausal women, postmenopausal women special diet, and premenopausal women respectively.

Calcium supplementation on body weight (Trowman, 2006)

Trial	Year	Baseline (weight kg)				Follow-up (weight kg)		Correlation between baseline and follow-up	
		Intervention		Control		Intervention	Control	Intervention	Control
		n	Mean (SD)	n	Mean (SD)	Mean (SD)	Mean (SD)	Correlation	Correlation
Chee	2003	91	56.1 (8.9)	82	57.2 (9.4)	*56.1* ***(8.9)***	*57.4* ***(9.4)***	0.957	0.962
Jensen	2001	25	94.6 (14.0)[a]	27	93.8 (14.0)[a]	89.0 (12.7)[a]	89.1 (14.7)[a]		
Lau	2001	95	56.9 (7.1)	90	58.9 (7.5)	*57.4* ***(7.1)***	*58.6* ***(7.5)***	0.933	0.935
Reid	2002	111	66.0 (10.0)	112	68.0 (11.0)	*65.7* ***(10.0)***	*67.9* ***(11.0)***	0.984	0.976
Shapses 1[c]	2004	17	84.1 (9.4)	19	89.4 (10.3)	*77.1* ***(9.4)***	*82.1* ***(10.3)***	0.880	0.868
Shapses 2[c]	2004	11	85.9 (9.2)	11	94.2 (15.7)	*79.2* ***(9.2)***	*86.6* ***(15.7)***	0.960	0.934
Shapses 3[c]	2004	18	93.7 (13.6)	24	93.5 (14.3)	*87.0* ***(13.6)***	*89.2* ***(14.3)***	0.918	0.970
Winters-Stone	2004	13	57.2 (4.9)	10	54.1 (7.2)	56.3 (4.3)	54.8 (7.2)		
Zemel	2004	11	99.8 (14.9)	10	103.1 (19.3)	*91.2* ***(14.9)***	*96.5* ***(19.3)***	0.937	0.910

a Calculated from the standard error
b Follow-up sample size ntrt = 24 and nctrl = 24
c Shapses *et al* (Shapses *et al*, 2004) report on three randomised controlled trials.
Trials 1, 2, and 3 include postmenopausal women, postmenopausal women special diet, and premenopausal women respectively.

Combining intervention estimates from CS only



An example: Calcium supplementation on body weight

- In this example, pooling SACS estimates is likely to produce a pooled estimate of intervention effect similar to pooling ANCOVA estimates.
- This occurs since even although there is baseline imbalance across the trials, the correlation between baseline weight and follow-up weight is likely to be large.
- What did the review authors do?
  - Pooled using estimates of intervention effect calculated from FV. When there were missing SDs at follow-up, they assumed baseline SDs (imputed SDs in 7 of 9 trials).
  - Also undertook a meta-regression adjusting for the baseline difference between groups.

Some final thoughts (1)

- Should we pre-specify a primary analysis in the protocol of the review?
  - For trials, it is common that we pre-specify the analysis in the trial protocol. This may include pre-specification of covariates which we will adjust for (e.g Senn 1989; Raab 2000).

OR …

- Should we undertake multiple analyses and draw conclusions from an interpretation based on the results from these analyses?

Some final thoughts (2)

- Option 2 (recreate the ANCOVA estimates) has advantages in terms of bias and precision, but it will generally involve imputing correlation coefficients (and perhaps SDs).
- Perhaps we could pre-specify an approach such as the following in review protocols:
  - Pool using results from only one analytical method (SAFV or SACS). In terms of bias, this will generally be reasonable. In terms of precision, this is not likely to provide an optimal solution.
    - The analytical method selected could be the one for which the summary statistics are most frequently reported. In the Trowman example, this would be CS.
  - If there is large baseline imbalance across the trials, a 'black belt' approach of recreating the ANCOVA estimates using a range of correlations, in separate re-analyses, could be undertaken.

References

- Overall JE, Magee KN. Directional baseline differences and type I error probabilities in randomized clinical trials.[comment]. *Journal of Biopharmaceutical Statistics* 1992;2(2):189-203.
- Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. Control Clin Trials 2000;21(4):330-42.
- Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989;8(4):467-75.
- Senn S. *Design and interpretation of clinical trials as seen by a statistician*. Chichester: Wiley & Sons, 1997.
- Trowman R, Dumville JC, Hahn S, Torgerson DJ. A systematic review of the effects of calcium supplementation on body weight. Br J Nutr 2006;95(6):1033-8.
- Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. J Clin Epidemiol 2007;60(12):1229-33.
- Wei L, Zhang J. Analysis of data with imbalance in the baseline outcome variable for randomized clinical trials. *Drug Information Journal* 2001;35:1201-1214.
- Williamson PR, Gamble C, Altman DG, Hutton JL: Outcome selection bias in meta-analysis. *Statistical Methods in Medical Research* 2005, 14:515-524.