Statistical Method for Meta-Analysis in Practical

Biostatistics Center
and
School of Public Health,
Taipei Medical University

Jin-Hua Chen

Outline

1. What is the Meta-analysis
2. Statistical method for meta-analysis
    1) Assess for heterogeneity
    2) Create summary measure (fixed effect or random effect model)
3. Example

1. What is the Meta-Analysis?

- Meta-analysis is the statistical procedure for combining data from multiple studies. When the treatment effect (or effect size) is consistent from one study to the next, meta-analysis can be used to identify this common effect. When the effect varies from one study to the next, meta-analysis may be used to identify the reason for the variation.

- In fact, both “Treatment effect” and “Effect size” can be used (and are used on this site) to refer to the identical set of indices - odds ratios, risk ratios, risk differences, standardized mean differences, raw mean differences, correlations, rate ratios, rate differences, and hazard ratios.
- What characterizes an effect size (and treatment effect) is that it looks at effects. Other meta-analyses do not look at effects but rather attempt to estimate the event rate or mean in one group at one time-point.

2. Statistical Method for Meta-Analysis

1) Accuracy of diagnostic test: sensitivity, specificity, sROC, Bivariate model, HSROC
2) Other statistics: OR, RR, mean, mean difference, proportion, proportion difference, HR, rate, correlation….

The standardized mean difference(continuous outcome) from Cochrane Handbook for Systematic Reviews of Interventions

- The **standardized mean difference(SMD)** is used as a summary statistic in meta-analysis when the studies all assess the same outcome but measure it in a variety of ways. In this circumstance it is necessary to standardize the results of the studies to a uniform scale before they can be combined.
- Hedges' *g* statistics: $g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}$, $\mathrm{SD}_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$$g_{corrected} = g \times J(n_1+n_2-2), \text{ where } J(n) = \frac{\Gamma(n/2)}{\sqrt{n/2}\,\Gamma\left((n-1)/2\right)}$$

$$SE(g) = \frac{n_1+n_2}{n_1 \times n_2} + \frac{g^2}{2(n_1+n_2)}, 95\%\ C.I.: g \pm 1.96 \times SE(g)$$

The standardized mean difference(continuous outcome) from Cochrane Handbook for Systematic Reviews of Interventions

- The Cohen's d statistics:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}}$$

is difference from Hedges' g statistics in SD_{pooled}.

- If the sample is large enough, the value of g and d estimates should be the same.
- When sample size is small, g is a biased estimator, but $g_{corrected}$ is an unbiased estimator.

2.1 Test of Homogeneity (同質性檢定)

$$H_0: g_1 = g_2 = \cdots = g_K$$

對於 K 個獨立研究結果的 SMD 是否相同的。

檢定統計量 Q 為(Cochran's *Q* statistic)

$$Q = \sum_{i=1}^{K} W_i \left(g_i - g \right)^2 \xrightarrow{\text{Under } H_0} \chi^2_{K-1}$$

在虛無假設為真的情況下, Q 服從自由度為K-1的卡方分配。

- 顯著水準為0.05或0.1

2.2 Create summary measure— Fixed Effect Model

- 可用I^2及同質性檢定來判斷研究間是否存在異質性。
  - I^2>50%:研究間異質性程度高; I^2<50%:研究間異質性程度低。
  - $I^2 = \frac{(Q-df)}{Q} \times 100\%$
- Fixed effect model所提的方法皆假設各獨立臨床試驗的測量值, 假設RR's, OR's, g's, 之間不存在異質性(Heterogeneity)。換言之，僅考慮 Within-study Variance而忽略Among-study Variance對試驗效果。
- 若各獨立臨床試驗的測量值之間確實存在異質性，則使用Random effect model來進行分析。

Combining effect size: g

$$\text{combining effect size } \mathrm{g} = \frac{\sum_{i=1}^{K} W_i g_i}{\sum_{i=1}^{K} W_i}$$

where g_i and W_i are the SMD and weighted in *i*th clinical trial.

We could obtain that

$$W_i = 1 / Var(g_i) = \frac{1}{\left(SE(g_i)\right)^2}.$$

or $W_i = f(n_i)$ is proportion of sample size.

Test of combining *g*

H_0: g = 0

$$\because Var(g) = \frac{\sum W_i^2 Var(g_i)}{\left(\sum W_i\right)^2} = \frac{1}{\sum W_i}$$

Under H_0 ,

$$Z_g = \frac{g - g_{under\ H_0}}{\sqrt{Var(g)}} \sim N(0,1)$$

95% CI for g is

$$\left(g_L\ ,\ g_U\right) = g \pm 1.96\ \sqrt{Var(g)}$$

2.2 Create summary measure– Random Effect Model

假設 $\theta_i = g_i$, $i = 1, \ldots, K$, 為第 i 個臨床試驗的效用(或評估之統計量)，滿足 $\theta_i \sim N(\mu, \tau^2)$。換言之，實驗效果

$$\theta_i = \mu + \delta_i,\ \delta_i \sim N(0\ ,\ \tau^2)$$

其中δ_i代表實驗效果的異質性源自於不同臨床試驗的群體特徵之差異(Ethnic Differences)或不同試驗Follow-up 的時間差異所造成，τ^2代表組間異質性的嚴重度。

根據同質性檢定統計量Q和加權W_i可以直接估計Among-study Variance, τ^2:

$$\hat{\tau}^2 = \max\left\{0\ ,\ \left[Q-(K-1)\right] \Big/ \left[\sum_{i=1}^{K} W_i - \left(\sum_{i=1}^{K} W_i^2 \Big/ \sum_{i=1}^{K} W_i\right)\right]\right\}$$

令 $W_i^* = 1 \Big/ \left(V_i + \hat{\tau}^2\right)$, 其中 $V_i = Var(\theta_i)$.

則Pooled實驗效果θ^*估計量為:

$$\theta^* = \sum_{i=1}^{K} W_i^* \theta_i \Big/ \sum_{i=1}^{K} W_i^* \quad \text{且} \quad SE\left(\theta^*\right) = 1 \Big/ \sqrt{\sum_{i=1}^{K} W_i^*}\ .$$

θ^* 的95%信賴區間為:

$$\left(\theta_L^* \ , \ \theta_U^*\right) = \theta^* \pm 1.96 \ SE(\theta^*)$$

θ^*有效性檢定的統計量為: $H_0 : \theta^* = 0$

$$U_\theta^* = \left(\sum_{i=1}^{K} W_i^* \theta_i\right)^2 \Big/ \sum_{i=1}^{K} W_i^* \sim \chi_1^2$$

Reject H_0, if $U_\theta^* > \chi_1^2(0.05) \approx 3.8415$

結論

1. $$W_i^* = 1\Big/\left(V_i + \hat{\tau}^2\right) \le 1\Big/\left(V_i\right) = W_i$$

$$SE\left(\theta^*\right) = 1\Big/\sqrt{\sum_{i=1}^{K} W_i^*} \ge 1\Big/\sqrt{\sum_{i=1}^{K} W_i} = SE(\theta)$$

2. 隨機效應模式估計(檢定)結果較保守。

3.各獨立臨床試驗間不存在異質性則採固定效應模式，反之，採隨機效應模式。

3. Example

Acupuncture treatment for pain: systematic review of randomised clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups

Matias Vested Madsen, Peter C Gøtzsche and Asbjørn Hróbjartsson

BMJ 2009;338;a3115
doi:10.1136/bmj.a3115

Topic collections Articles on similar topics can be found in the following collections

Pain (neurology) (3124 articles)
Pain (palliative care) (685 articles)
Pain (anaesthesia) (676 articles)
Complementary medicine (597 articles)
Drugs: musculoskeletal and joint diseases (2254 articles)
Internet (975 articles)

Abstract

- **Objectives** To study the analgesic effect of acupuncture and placebo acupuncture and to explore whether the type of the placebo acupuncture is associated with the estimated effect of acupuncture.
- **Design** Systematic review and meta-analysis of three armed randomized clinical trials.
- ……
- **Data** ..analysis Standardized mean differences from each trial were used to estimate the effect of acupuncture and placebo acupuncture.

Table 1 | Characteristics of trials

Trial	Clinical problem	Trial size—No randomised (No; % dropouts)	Blinding	Concealment of allocation	Pain scale	Treatment duration (No of sessions)*
Melchart[w1]	Tension headache	270 (30; 11%)	Patients	Centralised telephone randomisation	Rating scale (1-10)	8 weeks (12); evaluation at 12 weeks
Linde[w2]	Migraine	302 (20; 7%)	Patients	Centralised telephone randomisation	Rating scale (0-10)	8 weeks (12); evaluation at 12 weeks
Scharf[w8]	Osteoarthritis	1039 (57; 5%)	Patients	Central randomisation	WOMAC (0-10)	6 weeks (10); evaluation at 13 weeks
Witt[w7]	Osteoarthritis	300 (14; 5%;)	Patients	Centralised telephone randomisation	WOMAC (0-10)	8 weeks (12)
Foster[w9]	Osteoarthritis	352 (19; 5%)	Patients	Central telephone randomisation	WOMAC (0-10)	3 weeks† (6); evaluation at 6 weeks
Brinkhaus[w11]	Low back pain	301 (17; 6%)	Patients	Centralised telephone randomisation	VAS (0-100 mm)	8 weeks (12)
Molsberger[w10]	Low back pain	186 (12; 6%)	Patients	Central telephone randomisation	VAS (0-100 mm)	4 weeks (12)
Leibing[w12]	Low back pain	150 (36; 24%)	Patients	Unclear	VAS change (0-10 cm)	12 weeks (20)
Wang[w6]	Postoperative pain	101 (unclear)	Patients	Unclear	VAS (0-100 mm)	1 day (1)
Lin[w3]	Postoperative pain	100 (unclear)	Patients	Unclear	VAS (0-100 mm)	1 day (1)
Fanti[w5]	Colonoscopy	30 (unclear)	Unclear	Unclear	Rating scale (1-5)	1 day (1)
Sprott[w4]	Fibromyalgia	30 (unclear)	Unclear	Unclear	VAS (0-10)	3 weeks (6)
Kotani[w13]	Scar pain	70 (unclear)	Unclear	Sequentially sealed opaque envelopes	VAS (0-10 cm)	4 weeks (20)

VAS=visual analogue scale; WOMAC=Western Ontario and McMaster Universities pain subscale.
*Timing of evaluation is identical to treatment duration if not otherwise specified.
†Acupuncture and placebo acupuncture groups received three weeks of needling during six week standard care programme.

g_1=-0.318

W_i d_1=-0.316

Trial	Acupuncture Mean	SD	No	Placebo acupuncture Mean	SD	No	Weight (%)	Standardised mean difference (95% CI)
Brinkhaus[w11]	34.5	28.5	140	43.7	29.8	70	9.3	-0.32 (-0.61 to -0.03)
Fanti[w5]	2.0	1.3	10	2.8	1.6	10	1.0	-0.53 (-1.42 to 0.37)
Foster[w9]	6.38	4.1	113	5.98	4.3	115	11.5	0.09 (-0.16 to 0.35)
Kotani[w13]	2.5	2.3	23	5.6	1.2	23	0	-1.66 (-2.34 to -0.98)
Leibing[w12]	-2.7	2.2	35	-2.1	2.2	40	3.7	-0.27 (-0.73 to 0.19)
Lin[w3]	30.6	23.5	50	34.5	23.5	25	3.4	-0.16 (-0.65 to 0.32)
Linde[w2]	3.7	2.0	138	3.6	2.1	78	10.0	0.05 (-0.23 to 0.33)
Melchart[w1]	2.9	1.6	119	3.1	1.7	58	7.9	-0.12 (-0.44 to 0.19)
Molsberger[w10]	26.0	21.0	58	36.0	19.0	58	5.7	-0.50 (-0.87 to -0.13)
Scharf[w8]	3.0	2.3	315	3.3	2.4	358	33.7	-0.13 (-0.28 to 0.02)
Sprott[w4]	6.9	5.6	10	7.9	5.7	10	1.0	-0.17 (-1.05 to 0.71)
Wang[w6]	44.5	24.5	50	48.0	22.0	25	3.4	-0.15 (-0.63 to 0.33)
Witt[w7]	24.4	16.9	145	33.2	17.1	73	9.5	-0.52 (-0.80 to -0.23)
Total (95% CI)			1206			943	100.0	-0.17 (-0.26 to -0.08)

p8 Test for heterogeneity: χ^2=17.23, df=11, P=0.10, I^2=36%

p14 Test for overall effect: z=3.80, P<0.001

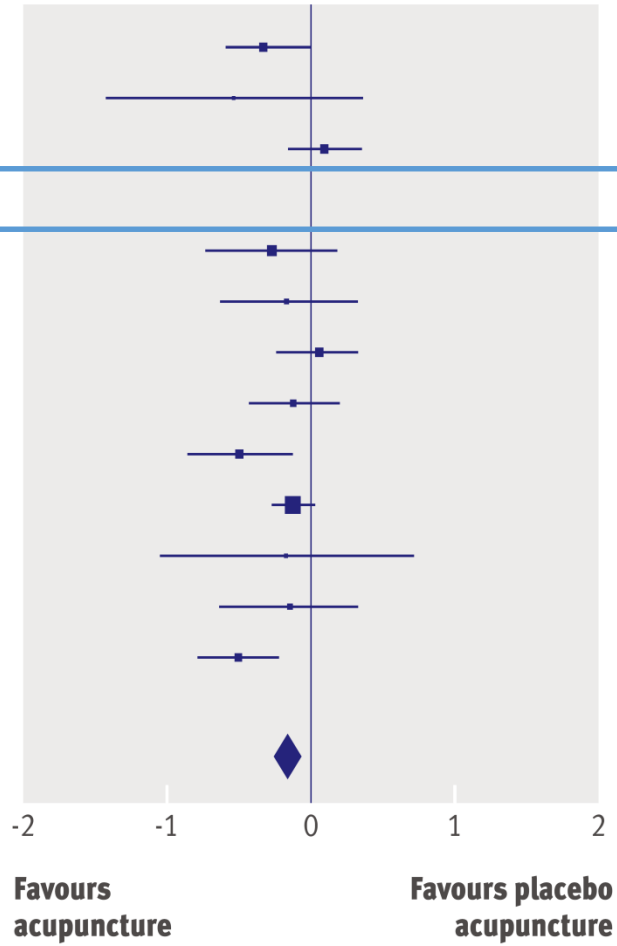

Fig 1 | Meta-analysis of acupuncture versus placebo acupuncture

Heterogeneity test

- Substantial heterogeneity was present in the comparison between acupuncture and placebo acupuncture ($P<0.001$, $I^2=66\%$).
- A trial by Kotani et al was a clear outlier—standardised mean difference −1.66 (95% confidence interval −2.34 to −0.98).
- We excluded this trial from all further analyses, after which the heterogeneity was substantially reduced ($P=0.10$, $I^2=36\%$).

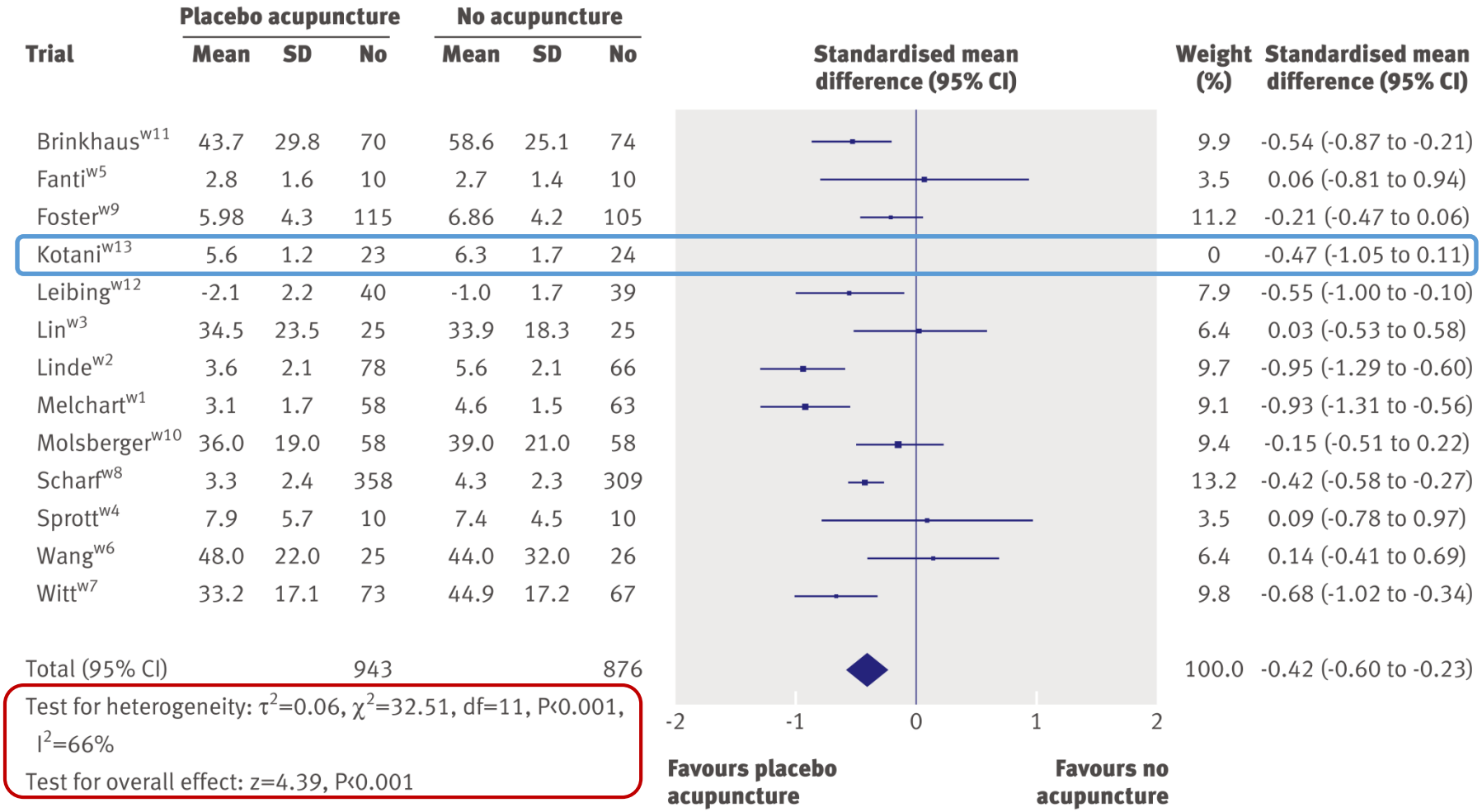

Fig 2 | **Meta-analysis of placebo acupuncture versus no acupuncture**

Secondary analyses

In two trials,[w3 w4] we could not define the authors' primary outcome, so the sensitivity analysis included 10 trials. In two trials,[w5 w12] our chosen outcome was the same as that of the authors (table 1). Substantial heterogeneity existed in the comparison between acupuncture and placebo acupuncture ($P<0.001$, $I^2=73\%$). The pooled standardised mean difference was −0.26 (−0.46 to −0.07) ($P< 0.001$). For the comparison of placebo acupuncture with no acupuncture, substantial heterogeneity was also present ($P<0.001$, $I^2=59\%$). The pooled standardised mean difference was −0.48 (−0.65 to −0.30) ($P=0.009$).

When we restricted the analysis to the seven trials with clearly concealed allocation, explicit blinding of patients, and dropout rate less than 15%, substantial heterogeneity existed in the comparison between acupuncture and placebo acupuncture ($P=0.01$, $I^2=63\%$). The pooled standardised mean difference was −0.19 (−0.35 to −0.02) ($P=0.03$). Heterogeneity was

Discussion

- We found a small difference between acupuncture and placebo acupuncture and a moderate difference between placebo acupuncture and no acupuncture. The effect of placebo acupuncture varied considerably.

Strengths and weaknesses

….

- Interpreting a standardised mean difference clinically may be challenging. On the basis of the mean standard deviation from the trials that had used visual analogue scales, the effect of acupuncture (standardised mean difference −0.17, −0.26 to −0.07) corresponds to a reduction of 4 (2 to 6)mm on a 100mm scale.

Conclusion

- We found a small analgesic effect of acupuncture that seems to lack clinical relevance and cannot be clearly distinguished from bias. Whether needling at acupuncture points, or at any site, reduces pain independently of the psychological impact of the treatment ritual is unclear.

The End