PharmaSUG 2012 - Paper SP04

APPLICATION OF META-ANALYSIS IN CLINICAL TRIALS

SAS® software for Meta-Analysis

Marina Komaroff, Noven Pharmaceuticals Inc., New York, NY

ABSTRACT

Drug development is a long and expensive process where each step should be carefully planned. During design stage of a clinical trial, the sample size calculation has to be performed based on a primary objective of the trial and target to achieve the desired power for detecting a clinically meaningful difference between test drug and standard/control at a fixed Type I Error rate. However, information about test drug in pilot studies is limited and most of the times statisticians do "the best guess" about effect size of a new drug that leads to wrong sample size calculation and then to failure of the trail.

Meta-analysis can help. Combining all existing information about the test drug, meta-analyses intent to give better estimation of effect size for a new drug that determines required sample. In current years, one of the obstacles was that meta-analysis is not an easy task and special software is usually required to perform it.

The goal of this paper is to demonstrate that concept of meta-analyses is apprehensible, and SAS® software can be used to perform meta-analysis on regular basis. User-friendly SAS macro calculates the effect size and determines what sample size is needed to reach the goal of clinical trial. For visual presentation of the result, macro generates accompanying forest plots of effect sizes and plot of anticipated sample size. For the sake of validation, the results from meta-analysis were compared with analysis of subject-level (pooled) data. The conclusions from two approaches came out the same. It demonstrates the validity and strength of meta-analysis.

This paper suggests that when data sets have been accumulated with ongoing research, the effect size calculated in meta-analysis can be treated as "best evidence" and should be taking in consideration while designing the next clinical trial that will lead to successful NDA submission.

INTRODUCTION

"Five clinical trials were conducted in my company for a new drug with one success and four failures…" sounds familiar for everyone who worked long enough in Pharmaceutical industry. "Hmm…" is your answer: "There is something wrong with expectations from a new drug, and sample size was probably too small. Why didn't you check the effect size of your company's new drug while accumulating data from study to study?" Let's do it. It is never too late.

Effect size is measured as standardized difference between two groups. Along with Confidence Limit Intervals (CLI), it gives information how different two samples are. In other words, effect size estimates the magnitude of treatment effect for a new drug. Effect size should be used to calculate sample size required to meet a p-value with level of significance α.

For this paper, data for ten studies were simulated using RAND function and normal distribution. Assumption was made that there are equal variances in the groups. Analysis for each study was done by PROC MIXED. Meta-analysis, and individual data analysis was performed for overall effect size calculations. The results from two approaches were compared. Sample size was proposed for a new clinical trial using estimated effect size of a new treatment. SAS® 9.2 was used for analyses.

DESCRIPTION

Assume that company is concentrated on developing a drug that reduces pain. Pain is measured in a continuous scale 0-100 mm. The more drop in score from baseline, the better is the treatment. The score assessment was performed at visits 1,2,3,4, and 5(end of study). The primary objective is to demonstrate the difference in the change from baseline between Active treatment and Placebo.

Ten studies have been conducted with different sample sizes. For meta-analysis, difference in means, standard deviation and 95% confidence limit intervals should be already in the final clinical study reports of the completed studies. If full data sets are available, the results of studies should be reproducible.

1) EACH STUDY EFFECT SIZE

The first step in meta-analysis calculation is to get the effect size estimation for each study. Calculation of effect size between two means can be done by three measures: Cohen's d, Hedges' g, and Glass' delta using formulas below.

SAS language is used in the text instead of mathematical formulas to be able to use highlighted lines right away into the program.

Cohen's d

```
     SIGMA_pooled=Stderr_Diff/sqrt((1/N_Placebo)+(1/N_Trt));
(d)  Effect_Size = Means_Diff/ SIGMA_pooled;

 Or  Effectsize_d= tValue*sqrt( ((N_Trt +N_Placebo)**2 ) /
                                    ((N_Trt*N_Placebo)*(N_Trt + N_Placebo -2)));
     Stderr_d=sqrt((N_Placebo+N_Trt)/(N_Placebo*N_Trt) +Effectsize_d^2/(2*(N_Placebo +
     N_Trt)));
```

Hedges' g

```
     sp=sqrt(((Stderr_Trt**2)*(N_Trt-1)+(Stderr_Placebo**2)*(N_Placebo-1))/(N_Trt +
     N_Placebo-2));
(g)  Effectsize_g=Mean_Diff/sp;
     Stderr_g=sqrt((N_Placebo+N_Trt)/(N_Placebo*N_Trt) +Effectsize_g^2/(2*(N_Placebo +
     N_Trt)));
```

Glass' delta

```
(Δ)  Effectsize_delta  =  Means_Diff/ Stderr_Placebo;
     Stderr_delta=sqrt((N_Placebo+N_Trt)/(N_Placebo*N_Trt) +
     Effectsize_delta^2/(2*(N_Placebo -1)));

      * 95% CLI for any effect size ***;
      Lowerbound_effect_size=effect_size -1.96 * stderr;
      Upperbound_effect_size=effect_size +1.96 * stderr;
```

2) EFFECT SIZE OVERALL

Calculation of the overall effect size (d) can be done for fixed and random effect models. Accessing the statistical heterogeneity will help to decide what model to use. If studies differ by sampling error only (homogeneous) than overall effect size from fixed-effects model is appropriate. If there is between-studies variability (heterogeneous), then random-effects model takes into account within-and between-studies variability, and overall effect size from this model should be used.

Fixed-effect model

The weight of the individual study results is proportional to the square of standard errors:

```
     Wi=1/(stderr_effectsize^2);
```

The weighted average effect size (Theta) will be:

```
     Theta=sum(effectsize_i * Wi)/sum(Wi) ;
     Std_Theta=sqrt(1/(sumWi));
     Upper_Theta=Theta+1.96*Std_Theta;
     Lower_Theta=Theat-1.96*Std_Theta;
```

where i-is the number of the study.

Assessing Heterogeneity of studies

Cochran's Q statistics can give estimation of true heterogeneity among the studies effects.

```
     Q=sum{Wi*(effectsize_i - Theta)^2};
```

Q-statistic has chi-square distribution with k-1 degree of freedom (k is the number of effect sizes in the sample). High values of Q (above the critical point for a given significance level α) enable us to reject the null hypothesis about homogeneity and conclude that there is statistically significant variation between studies.
If the value of Q is unusually large then there is substantial heterogeneity, because there is more variation among the studies than expected just by looking at the individual standard errors. If there is no heterogeneity, then Q should be approximately equal to k-1.
Knowing Q, the estimator of between-study variance tau can be calculated using Method of Moments. [4]

```
If Q > k-1 then tau^2= (Q-(k-1)) / C;
If Q ≤ k-1 then tau^2= 0 ;
```

Where `C = sum (Wi) – (sum (Wi^2))/ (sum(Wi));`

The extent of heterogeneity can be estimated by I^2 index by comparing the Q value with its expected value assuming homogeneity. It measures proportion of inconsistency in individual studies, and values close to 100% imply very high degrees of heterogeneity. I^2 equal to 25%, 50%, and 75% represent low, moderate, and high heterogeneity accordingly.

```
If Q > k-1 then I^2=100 * (Q-(k-1)) /Q;
If Q ≤ k-1 then I^2= 0;
```

Random-effect model

Random effects model can be considered as the case of fixed effects model with addition of between-studies variability. The weights will be calculated by adding tau^2 in the formula.

```
Wi_random=1/( tau^2  + stderr_effectsize^2 );
```

The weighted average effect size (Theta) will be:

```
Theta_random=sum(effectsize_i * Wi_random)/sum(Wi_random) ;
where i-is the number of the study.
```

Use the following formulas to calculate standard deviation for Thetas along with 95% CL intervals.

```
Std_Theta_random=sqrt(1/sum(Wi));
Lower_Theta_random=Theta_random – 1.96*std_Theta_random;
Upper_Theta_random=Theta_random + 1.96*std_Theta_random;
```

Confidence intervals for random effects models are usually wider than a fixed effects model. It happens because the estimated study heterogeneity adds uncertainty to the confidence interval calculations.

Sample Size

The sample size required for calculated Effect size in units and 2-tail p-value with level of significance α will be calculated by formula: N= (Z * GroupSD/Effect_size)^2 ; where Z value=Mean_Diff/GroupSE.

EXAMPLE

In this paper example, our new drug is supposed to reduce a headache. Subjects complain about the pain and report its level (scale: 0 to 100) every day in the diary. Study drug is compared to placebo at visit 1, 2, 3, 4, and 5(the end of clinical trial). The change from baseline in the level of pain demonstrates that study drug had significantly high (let's hope) reduction in pain compared to placebo subjects at the end of study (Visit 5) at level of significance α=0.05. LS Means were used in this example, because of the adjustment for baseline in the model. Data for 10 studies were simulated where each study has two arms with different number of subjects.

Study	N0 (placebo)	N1 (Active)	Total
1	25	27	**52**
2	50	53	**103**
3	75	79	**154**
4	100	105	**205**
5	125	131	**256**
6	150	157	**307**
7	175	183	**358**
8	200	209	**409**
9	225	235	**460**
10	259	261	**511**

1) Difference in LS Means (95% CI) is calculated for each study. The results are summarized in Table 1

Table 1: LS Means* difference along with 95% CL Intervals between Active and Placebo

Study	Visit	N Placebo	LS Mean Placebo	STD ERR Placebo	N Active	LS Mean Active	STD ERR Active	LS Means Difference	STD ERR Diff	t-Value	95% CL interval Lower	95% CL interval Upper
1	5	25	-0.44	5.24	27	6.85	5.04	**7.29**	7.27	-1.00	**-7.02**	**21.61**
2	5	50	9.83	3.56	53	10.97	3.46	**1.15**	4.96	-0.23	**-8.60**	**10.90**
3	5	75	3.26	3.09	79	8.87	3.01	**5.61**	4.32	-1.30	**-2.86**	**14.08**
4	5	100	0.39	2.47	105	6.94	2.41	**6.55**	3.45	-1.90	**-0.22**	**13.31**
5	5	125	4.21	2.14	131	6.95	2.09	**2.74**	2.99	-0.92	**-3.13**	**8.61**
6	5	150	0.40	2.04	157	3.07	2.00	**2.67**	2.85	-0.94	**-2.93**	**8.27**
7	5	175	0.11	1.86	183	12.18	1.81	**12.07**	2.59	-4.65	**6.98**	**17.16**
8	5	200	4.38	1.76	209	8.10	1.72	**3.72**	2.46	-1.51	**-1.10**	**8.55**
9	5	225	2.45	3.04	235	9.14	3.02	**6.69**	2.34	-2.86	**2.10**	**11.27**
10	5	250	1.79	1.60	261	9.84	1.57	**8.06**	2.24	-3.60	**3.66**	**12.45**

*Adjusted for baseline in the model

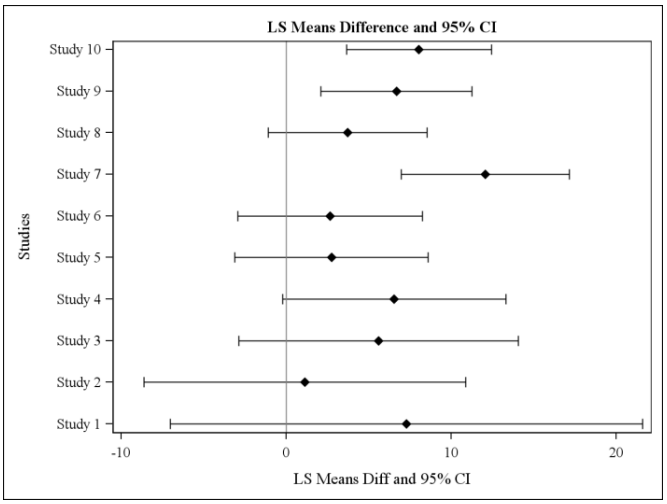
Any effect size can be calculated using formulas from the description section. This example use Cohen's effect size. Table 2 presents Cohen's effect sizes together with 95% CL intervals.

Table 2: Cohen's (d) Effect Size

Study	Visit	LS Mean Diff	STD ERR Diff	Sigma Pooled	Effect size (d)	Std Err (d)	Lower Bound (d)	Upper Bound (d)
1	5	**7.29**	7.27	26.18	**0.28**	0.28	**-0.27**	**0.83**
2	5	**1.15**	4.96	25.17	**0.05**	0.20	**-0.34**	**0.43**
3	5	**5.61**	4.32	26.77	**0.21**	0.16	**-0.11**	**0.53**
4	5	**6.55**	3.45	24.67	**0.27**	0.14	**-0.01**	**0.54**
5	5	**2.74**	2.99	23.93	**0.11**	0.13	**-0.13**	**0.36**
6	5	**2.67**	2.85	25.00	**0.11**	0.11	**-0.12**	**0.33**
7	5	**12.07**	2.59	24.54	**0.49**	0.11	**0.28**	**0.70**
8	5	**3.72**	2.46	24.87	**0.15**	0.10	**-0.04**	**0.34**
9	5	**6.69**	2.34	25.06	**0.27**	0.09	**0.08**	**0.45**
10	5	**8.06**	2.24	25.31	**0.32**	0.09	**0.14**	**0.49**

Figure 1 depicts LS Mean Difference together with 95% CL intervals for all 10 studies using original units. It can be seen, that there is a difference between active treatment and placebo to reduce pain in favor of a new drug; but most studies (#1-6, and 8) didn't demonstrate the significant reduction.

Figure 1: LS Means Difference (95% CI) for 10 Studies



Next step is calculation of overall effect size for 10 studies.

2) Fixed model effect sizes for LS means differences and Cohen's effect sizes (d) were calculated using SAS lines from the description section (Table 3).

Table 3: Effect Size for Ten Studies (Fixed Effect Model)

Visit	Effect Size Θo (original units)	Std Err Θo (original units)	Lower Bound Θo (original units)	Upper Bound Θo (original units)	Effect Size Θ	Std Err Θ	Lower Bound Θ	Upper Bound Θ
Visit 5	6.13	0.94	4.28	7.97	0.24	0.0379	0.17	0.32

3) Test for Homogeneity generates Q and I^2 statistics, as well as Tau^2 for LS means differences and Cohen's effect sizes (d). Q-statistic values that close to k-1 suggest homogeneity across studies (Table 4). I2 is equals to 26% that indicates the low level of heterogeneity among the studies.

Table 4: Homogeneity Tests for Ten Studies

Visit	C	Cd	K	Q	Qd	Tau2	I2	Taud2	Id2
5	608.42	0.98	10	10.60	10.80	0.0026	0.2628	1.62	162.48

4) Knowing Tau^2, the overall effect size for LS means difference and Cohen's effect size (d), the results from random-effects model can be easily calculated (Table 5). Since there is homogeneity across studies, overall effect sizes for fixed and random models are very close.

Table 5: Effect Size for Ten Studies (Random Effect Model)

Visit	THETAd Random	STD THETAd Random	Lower THETAd Random	Upper THETAd Random	THETA Random	STD THETA Random	Lower THETA Random	Upper THETA Random
Visit 5	6.05	1.04	4.01	8.10	0.24	0.0420	0.16	0.32

5) Individual Data Analysis was performed to compare the results.
The hardest part for individual data analyses is combining data together especially for studies that completed long time ago and data might have been collected in different units and formats.
There was no problem for this example because data were simulated. PROC MIXED gave the desirable results using study as random variable (Table 6). It is easy to see that estimation of effect sizes from meta-analysis (Table 4 and Table 5) and individual data analysis (Table 6) are very close.

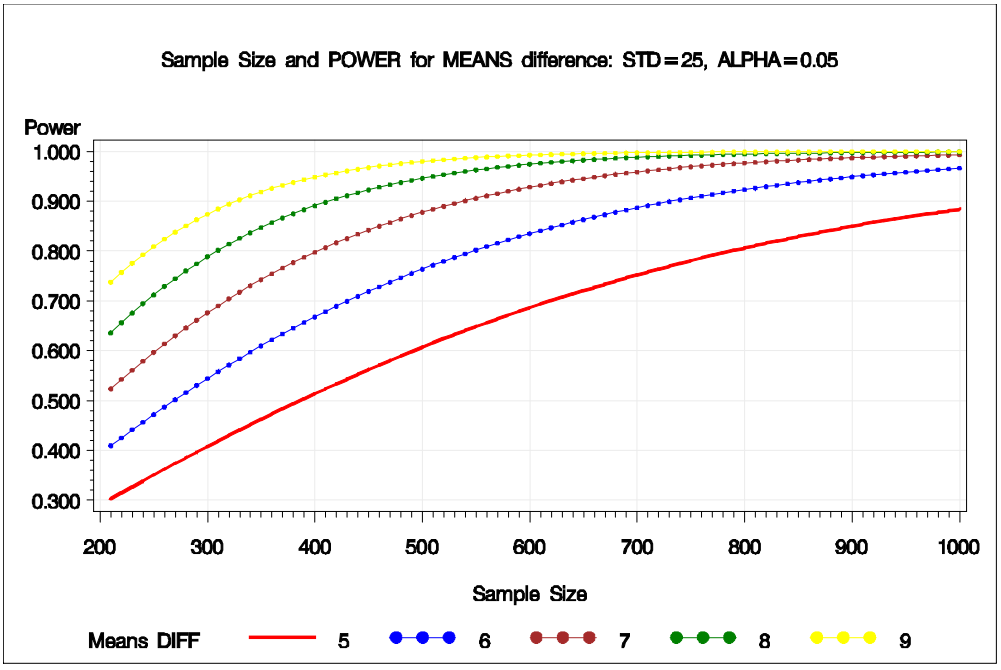
Table 6: Individual Data Analysis: LS Mean Difference along with 95% CL interval (Random Model)

Visit	LS Mean Difference	Lower 95%	Upper 95%	Effect size (d)	Std Err (d)	Lower Bound (d)	Upper Bound (d)
5	6.13	4.29	7.98	0.25	0.04	0.17	0.32

6) Sample Size calculation.

Overall effect size from random-effects model was used as difference in means equals to 6.00. The group sample size of 550 (275 per arm) will achieve 80% power to detect means difference of 6.0 between the two groups with known standard deviations of 25.0 for both groups, and with a significance level (alpha) of 0.05 using a two-sided two-sample t-test (Figure 2: Sample Size and Power).
This can explain why studies #1-6 and 8 were not successful. Their sample sizes were too small to detect small effect size of a new drug.

Figure 2: Sample Size and Power



VALIDATION

Validation of the macro was performed by comparing the results using validated Comprehensive Meta-Analysis software (Copyright ©2006 Biostat, Inc.) on the same data sets. Results matched.

LIMITATION

Macro was written for comparing means of two samples. Another step should be taken for other choices.

CONCLUSION

It is very important to estimate the effect size of a new drug correctly. When data sets have been accumulated with ongoing research, meta-analysis can be very useful. The effect size calculated in meta-analysis can be treated as "best evidence" and should be taking in consideration while designing next clinical trial and performing a sample size calculation.

Example in this paper demonstrated that results from meta-analysis are very similar to individual data analysis. Individual data analysis requires more time and resources to pool data together from various studies. Moreover, very often subject-level data is not available at all. Meta-Analysis can be performed with less effort using SAS® and help drug companies with proper effect size estimation that leads to successful NDA submission.

ACKNOLEDGEMENT

I want to thank my supervisor Sailaja Bhaskar, PhD for motivation and valuable support.

REFERENCES

1. **"Computing Effect Size Measures with ViSta - The Visual Statistics System"** Rubén Daniel Ledesma, Guillermo Macbeth, Nuria Cortada de Kohan. Tutorials in Quantitative Methods for Psychology 2009, Vol. 5(1), p. 25-34.
2. **"Evidence-based sample size calculations based upon updated meta-analysis"** Alexander J. Sutton, Nicola J. Cooper, David R. Jones, Paul C. Lambert, John R. Thompson and Keith R. Abrams. Statist. Med. 2007; 26:2479–2500
3. **"Meta-analysis Notes"** DeCoster, J. (2004) from http://www.stat-help.com/notes.html
4. **"Assessing Heterogeneity in Meta-Analysis: Q Statistic or I2 Index?"** Tania B. Huedo-Medina, Julio Sa´nchez-Meca, and Fulgencio Marı´n-Martı´nez, Juan Botella; Psychological Methods 2006, Vol. 11, No. 2, 193–206.
5. **" How to calculate effect sizes from published research: A simplified methodology"** Will Thalheimer Samantha Cook; A Work-Learning Research Publication; Copyright 2002 by Will Thalheimer

CONTACT INFORMATION

Marina Komaroff, MPH, MS, DrPH Student
Senior Manager, Biostatistics,
Clinical/Regulatory Operations
Noven Pharmaceuticals, Inc.
Empire State Building
350 Fifth Avenue, 37th Floor
New York, NY 10118
Tel: 212 299 4202
Email: Mkomaroff@noven.com