

Some Basic Points Concerning Meta-Analysis and Clinical Trials

by

**Daryl S. Paulson, Ph.D
President and Chief Executive Officer
BioScience Laboratories**

September 9, 2011



Some Basic Points Concerning Meta-Analysis and Clinical Trials

In the past, it has been very useful to perform statistical analyses on studies to evaluate products for their antimicrobial capabilities. If done correctly, this can determine if the product is effective or not.

But over the years, there have been multiple studies completed, evaluating the same product in the same general ways. What does one do when one study says a product is effective, but another study says it is not effective? That is the quandary we are now in.

A statistical methodology that has come of age during the computer revolution allows one to evaluate studies done at different times of the year, by different technicians on different test subjects, and combine these factors into one study – a Meta-Analysis (1, 2). Meta-Analysis has been developing at a quick pace. But what is it specifically? Meta-Analysis uses the results gained from other studies as its data points. It takes these points (results of other statistical studies) and combines them into a summary statistic (11). The Meta-analysis data can be the weighted p -values or the weighted difference between baseline and post-application sample – the weighted effect size. The effect sizes will be used in this paper, because they are more accurate than the p -values (7).

Meta-Analysis for Continuous Data

There are many things that Meta-Analysis can do, but our focus will be on continuous data instead of binary or correlational data. For example, suppose we are interested in a difference between several categories: 1) two different products (i.e., two chlorhexidine gluconate formulations), or 2) a test and a control product, or 3) the baseline and a sample time value of one product. Any combination is suitable. In this example, we will look at a baseline and a sample time.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

The difference between the baseline and sample time for each study is calculated by the D value.

$$D = \frac{\bar{x}_{b_1} - \bar{x}_{sample\ time}}{s_{pooled}}$$

where:

D = the difference (baseline minus sample time) divided by the s_{pooled} value

$\bar{x}_{sample\ time}$ = the \log_{10} count average of the sample time

\bar{x}_{b_1} = the \log_{10} colony count average of the baseline

This study involved the baseline and the wash time colony counts on the same subject (i.e., left versus right hand). The hands were selected for baseline and wash time according to the randomization schedule. This is a paired test in which a pooled standard deviation is used:

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where:

s_{pooled} = standard deviation of baseline and sample time combined

n_1 = sample size of the application data

s_1^2 = variance of the application data

n_2 = sample size of the baseline data

s_2^2 = variance of the baseline data

The variance of D is given by:

$$V_D^2 = \frac{n_1 + n_2}{n_1 n_2} s_{pooled}^2$$

and the standard deviation by:

$$\sqrt{V_D^2}$$

Some Basic Points Concerning Meta-Analysis and Clinical Trials

Hedge's g Statistic

The Hedges g statistic is computed as:

$$g = J \times D$$

where:

$$J = 1 - \frac{3}{4df - 1} \quad (df = \text{degrees of freedom})$$

D = from above (mean difference divided by the pooled standard deviation)

The statistics are rather complicated and usually performed by a computer software system, so they will not be discussed further. For a background of how different statistical programs may be calculated, one can review the procedures (1, 3).

Fixed or Random Effects

Knowledge of these effects is very important in Meta-Analysis, for using fixed versus random effects changes the confidence intervals, as well as the grand total, often dramatically. The definitions are different from the normal parametric statistics in “fixed” and “random” effects. In parametric statistics, if we are given the parameters (e.g., three specific products and 1- and 10-minute sample times) and do not choose them from all products and time frames, they are fixed effects. Before the study begins, if the products used are drawn at random from among all products available, or if random sample times are selected, it is a random effects model.

In Meta-Analysis, this is changed (9). For the fixed effects model, it is assumed that there is one true effect for all studies involved. In other words, a drug's effect or the reduction in colony counts of bacteria are the same.

For the random effects, the true readings are not the same for all the studies. For example, the baseline bacterial average counts on the hand surfaces may be decidedly different

Some Basic Points Concerning Meta-Analysis and Clinical Trials

among races, sexes, different locations of the study, and times of the year. Hence, there may be different baseline average counts among the groups. This would best be estimated by random effects.

In summary, for the fixed effects:

- There is one true effect size for all studies
- All the different effects are really sampling errors
- Weights are assigned high for studies that have high sample sizes
- The small sample size studies get smaller weights and the larger studies higher weights
- Weights of study = $\frac{1}{V_y}$, where $V_y = s^2 = \text{variance}$

Random Effects:

- There are different true effects sizes, depending upon the study
- The different effects sizes are real, not sampling errors
- One does not assign higher weights to studies with larger sample sizes
- The small sample size studies do not get smaller weights
- Weights of study = $\frac{1}{V_y^*}$, where $V_y^* = s^2 + \tau^2$, $s^2 = \text{variance}$, $\tau^2 = \text{within-subject variance}$

Which Model Should be Used?

The main effects (baseline minus wash time) should be random effects for these studies, because they are performed at different times by different people using separate subjects. But it makes sense to use fixed effects if two conditions are met. First, if one believes that all the

Some Basic Points Concerning Meta-Analysis and Clinical Trials

studies included are identical, and second, if our goal is to compare a common effects size for identical populations (3).

These two facts are not commonly found in antimicrobial studies. First, for studies where media is put on the hands, the initial population probably will vary, providing different baselines among the different studies. For studies that use the normal population, it will depend on the time of year, humidity, and the subjects for their baseline bacterial counts.

The second question is will the sample size be consistent among studies? Some studies have $n = 5$ subjects, and other studies have $n > 100$; so they will vary. To be safe, use the random effects model (9).

This is not all that is happening with this study, as will be discovered later when the subgroups – the application times (30 seconds and 1 minute) – were put into the model. These applications times would be consistent no matter what product was tested, by whom, or in what subgroup. This will make the subgroup effect a “fixed effect.” We will discuss this later.

Importance of Selecting All Studies, Not Just the “Good” Ones

In Meta-Analysis, this is the most important point. It is critical to select all the studies that one can find for the evaluation being performed (4). When beginning a meta-analysis, one must have an inclusion/exclusion criteria list for the studies. For example, the inclusion criteria designated all studies that used the FDA handwash guidelines (4) for surgical scrubs. Notice that these items are not the way “this test is supposed to be run,” but the way it was designed. The exclusion criteria consisted of studies for which the data generation was not understood, types of studies using guidelines different from those of the FDA, and studies that were not performed in a randomized manner. These two areas will require much time for the precise selection of studies to be used (5).

Some Basic Points Concerning Meta-Analysis and Clinical Trials

Table 1 has a series of studies included.

Table 1. Study Data

	Study Name	Baseline			Application of Product		
		Group A Mean	Group A Std Dev	Group A Sample Size	Group B Mean	Group B Std Dev	Group B Sample Size
1	12	6.800	0.340	30	3.560	0.265	30
2	13	7.400	0.389	50	3.890	0.367	50
3	24	6.880	0.678	38	4.789	0.567	38
4	45	7.900	0.564	20	4.890	0.452	20
5	67	5.890	0.780	10	3.870	0.959	10
6	71	8.900	0.561	62	6.900	0.780	62
7	26	7.520	0.294	75	3.190	0.379	75
8	48	6.300	0.593	16	4.870	0.362	16

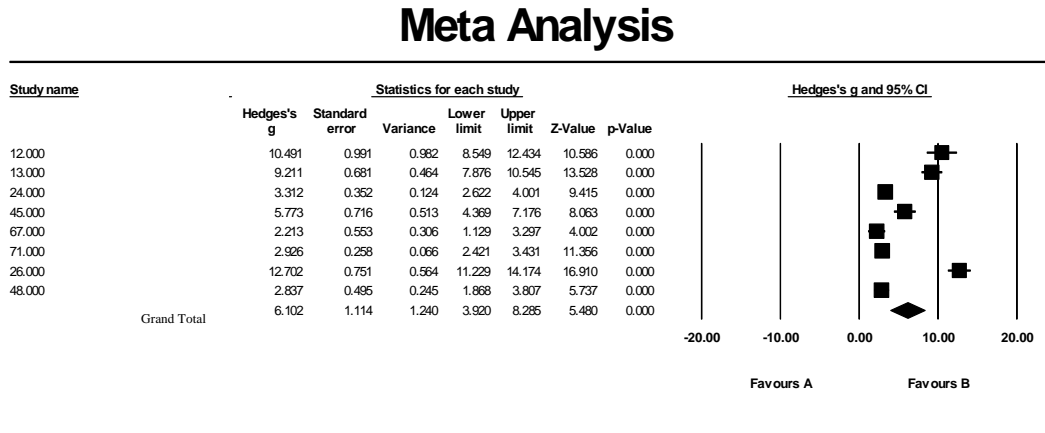
The eight studies in this meta-analysis fit the inclusion/exclusion criteria.

Meta-Analysis

The preliminary meta-analysis is displayed in Figure 1, where each of these studies went through an analysis and got a final or grand total score (bottom line). The 95% confidence intervals are also given, with a probability value.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

Figure 1. Meta-Analysis



Meta Analysis

Each of these eight studies was significant ($p \leq 0.000$). Notice at the very bottom of the graph, the diamond represents the average of all eight of the tests. The value of the Hedges $g = 6.102$ ($p < 0.000$). The values were synthesized into one value for the entire meta-analysis.

Looking at the graph portion (right-hand side) of Figure 1, there seems to be significant difference among the groups, even though this is a random effects study, in which variation among studies is expected. The Hedge's g values do not seem to be homogenous, but instead, heterogeneous. The model was changed to a fixed effects model for a moment. We can now test them.

H_0 : All groups are homogenous.

H_A : The groups are different (heterogeneous).

We find for the heterogeneity test that the Q value is 278.472, with 7 degrees of freedom.

$$278.472 - 7 = 271.472$$

Some Basic Points Concerning Meta-Analysis and Clinical Trials

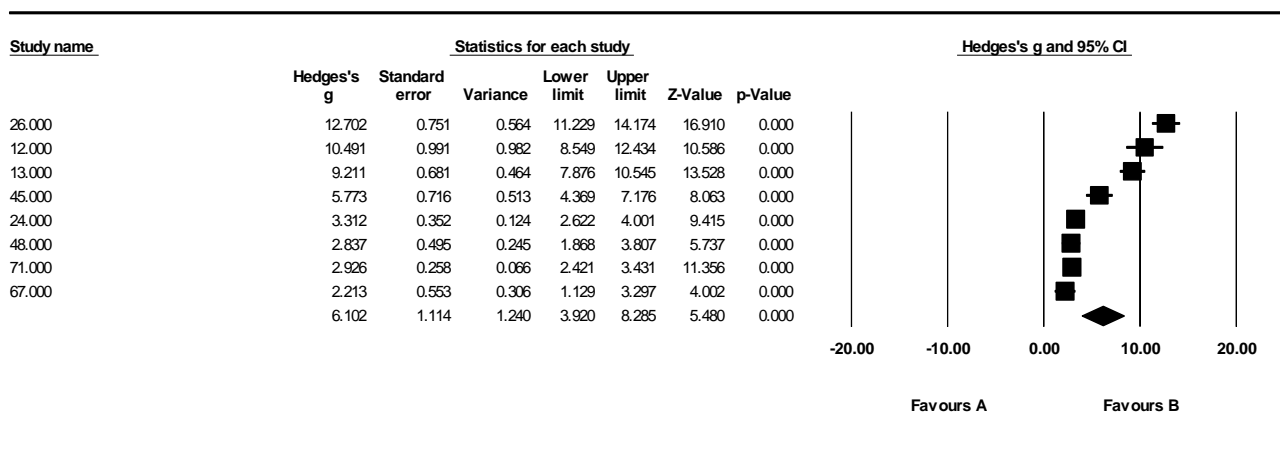
Using the Chi Square test and checking 271.472, we see that it is significant ($p < 0.000$). The results are not homogenous.¹

Look at the data (Figure 1), we can see that Study numbers 12, 13, and 26 are different from the other studies in that they appear to be more effective. But why?

Then, the data were rearranged from high to low and reviewed (Figure 2).

Figure 2. Meta-Analysis Table (High to Low Arrangement)

Meta Analysis



Meta Analysis

Going back to the original studies and seeing if the products were different or if the application times were larger, we notice that there was a difference. There were two product application times, 1 minute and 30 seconds. We failed to see this at first but found it by assuming the model was fixed effects and looking for heterogeneity. A subgroup (time of wash) was then put into the model. We dare not go any further, because this is a random assignment model, and errors are expected. The model was then changed back to random effects.

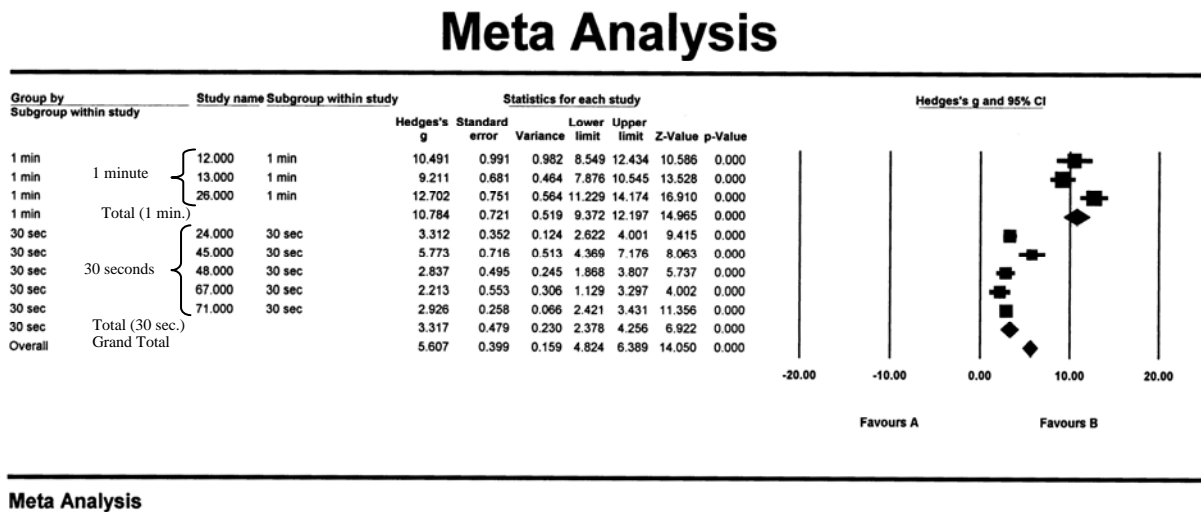
¹ Note: A test for homogeneity cannot be conducted on a random effects model, because we assume that the values differ. This study was temporarily assigned to a fixed effects category to see if there was heterogeneity. The test data is Q value; degrees of freedom and the Chi Square test were not shown.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

The Final Model

So we have a random effects general study model, which is embedded with subgrouped times that are fixed effects. This provides a “mixed effects” model. The model selected is also 1) an analysis across levels of the two subgroups, and 2) comparing the effects of these subgroups (10). This study had a common variance that was pooled. Figure 3 presents these data.

Figure 3. Meta-Analysis: Two Time Points (30 second and 1 minute)



There are two sub-analyses going on in this table. The first (summarized by the first diamond) is for a one-minute application, which gave a Hedges g of 10.784 ($p < 0.000$). This is highly significant. There is also a 30-second application. This was not as effective as the one-minute application, but it was very effective from baseline to application time. It is summarized by the second diamond, a Hedges g statistic was 3.317 ($p < 0.000$) that is very effective. The 30-second and one-minute applications were combined into an overall Hedges g statistic (the third diamond), which was 5.607 ($p < 0.000$).

The product kills at 30 seconds, but it kills many more bacteria if applied for one minute. These are statistically significant results, but some say that clinical significance would need to be

Some Basic Points Concerning Meta-Analysis and Clinical Trials

evaluated using the two applications in the hospital setting. This would not be done, due to ethical reasons, so it is best to refer to the statistical significance (4).

Looking for Bias

If the studies were all-inclusive, then there would be no need to look for bias. But we do not know if they are. There are two situations where bias may exist. The first is that studies that oppose one's beliefs are eliminated. For example, a person might choose only the studies that show that their product was superior to others (3). To this end, significant studies were evaluated, and insignificant ones were not. This is a major problem (1, 2, 3). The second case is that contradictory studies may not have been published. For example, very small studies or studies that show no effects are rarely published (10). It should also be recognized that studies that make a large difference are more likely to be published than studies that do not. This presents potential biases.

Because bias cannot be avoided with certainty, we can assess its potential by formulating a few questions:

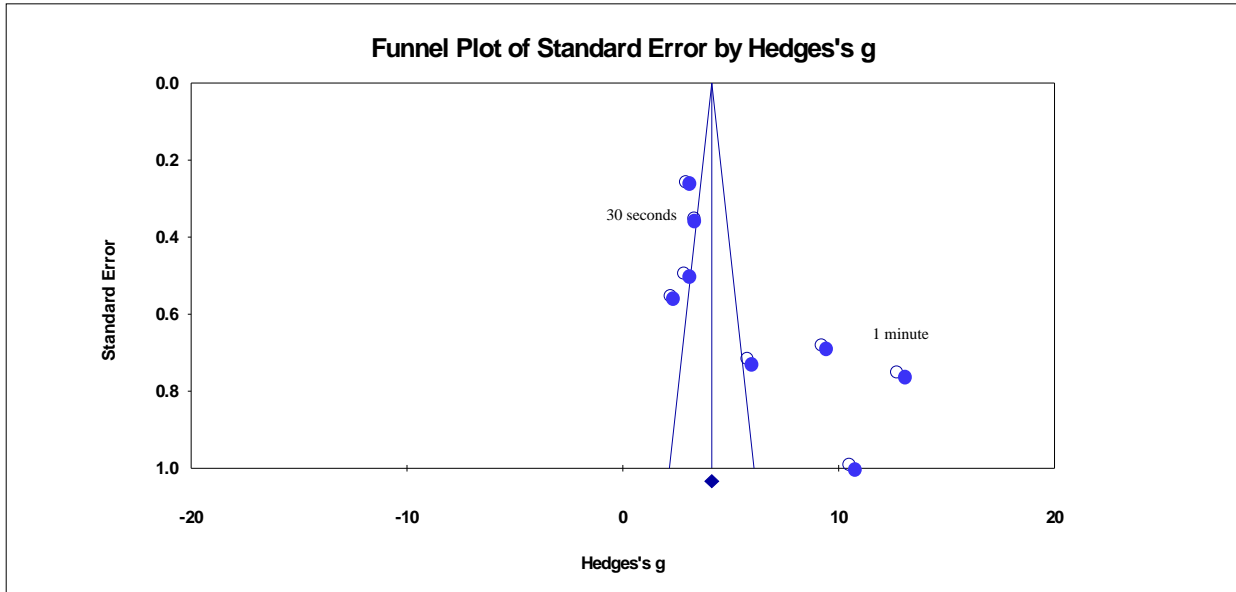
1. Is there evidence of bias?
2. Is it possible that the entire effect is due to an artifact of bias?
3. How much impact of bias is present?

The Cochrane Collaboration (6) has published the results of over 3700 meta-analyses, but none are relevant to evaluating topical antimicrobials the way the FDA evaluates them. There were several studies comparing the incidence of disease relative to hand-washing, but this is not how this meta-analysis was designed.

A good place to begin is with a funnel plot (Figure 4). We can see that it is composed of standard error versus the Hedges g statistic. The smaller the study, the larger the standard error.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

Figure 4. Funnel Plot



In this study, we have a problem: two different time frames. We could separate the 30-second and 1-minute times and perform two different funnel plots. However, we would have limited the studies to only three data points for the 1-minute time and only five for the 30-second time. This is not enough data to detect a bias if one exists, so we will leave the study undivided.

We already know that the three highest Hedges' g studies were done at one-minute application times instead of 30-second. We can see them pulling the Hedges g to the right. Using the Duval and Tweedie's trim and fill statistic (7) (Figure 5), we take the mirror image of the most weighted aspect of the studies, so they are neutralized.

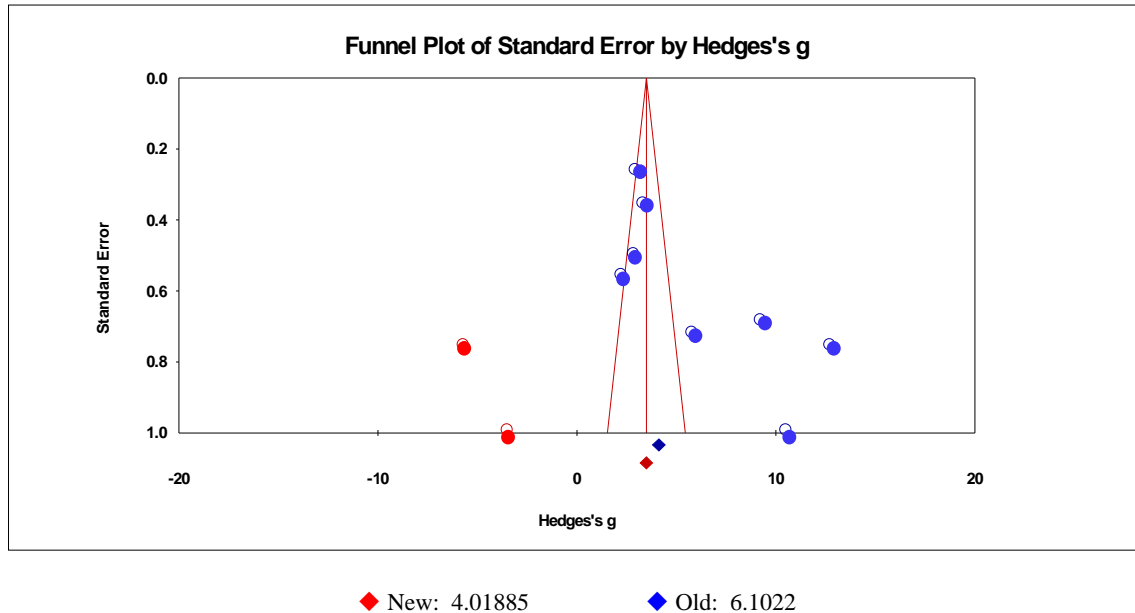
Figure 5. Duval and Tweedie's Trim and Fill

Random Effects					
	Studies	Point			
	Trimmed	Estimate	Lower Limit	Upper Limit	Q Value
Observed values		6.10220	3.91971	8.28468	278.47205
Adjusted Values	2	4.01885	1.56302	6.47467	493.47963

Some Basic Points Concerning Meta-Analysis and Clinical Trials

Figure 6 shows what happened. Two values were placed as their mirror images to see what would happen.

Figure 6. Funnel Plot of Standard Error by Hedges g



The average grand total point went from 6.1022 to 4.01885. The upper and lower confidence intervals are also presented. This is not actually true, for we found three studies that were performed at one-minute application times and the other was performed at 30 seconds. If we ignore this for a moment, we see the study dropped to the left. Figure 6 is a funnel plot of these data. We see that the mirror image of the highest two values on the right is also placed on the left, and the Hedges g statistic is recomputed. That average effect is 4.01885 for both time frames. At worst case, the study results would continue to be significant, even though the average effect has moved to the left.

A Different Kind of Analysis

One must discover the best way to report the data relative to the readers' comprehension. For example, what does one do when there is a baseline, Wash 1 (immediate sample) and Wash 10 (sample after 10 washes) when evaluating a healthcare handwash?

Some Basic Points Concerning Meta-Analysis and Clinical Trials

The researcher must determine the best way to present the data. Perhaps it is in a combined comparison within one study. For example, we could subtract Wash 1 and Wash 10 from baseline. This will eliminate the baselines from beginning at different values. In the same study, we can follow both times. The math is more difficult, for we have a variance that is probably small for the first wash but gets much larger after the tenth consecutive wash. We must find the correlation coefficient and take it into account in calculating the variance. Or we can do two meta-analyses, one for Wash 1 and one for Wash 10. The key questions will be “what will be easier for readers to understand?” and “How can data best be presented in an unbiased manner?”

Conclusion

Meta-analysis is now coming of age. It is a strong, emergent statistical methodology, which allows one to test a product's ability when various studies are involved and integrate the results in an easily understandable data display.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

REFERENCES

1. Cooper H., Hedges, L. V., & Valentine, J. G. *The Handbook of Research Synthesis and Meta-Analysis (2nd Edition)*. Russell Sage Foundation, New York. 2009.
2. Egger, M. Smith, G. D., & Altman, D. B. *Systematic Reviews in Healthcare: Meta-Analysis in Context*. BMA House, London. 2009.
3. Borenstein, M. Hedges, L. V. Higgins, J. P. T., & Rothstein, H. R. *Introduction to Meta-Analysis*. Wiley & Sons, West Sussex, UK. 2009.
4. Federal Register. 21 Code of Federal Regulations, parts 333 and 369. Tentative Final Monograph for Health Care Antiseptic Drug Products. Friday, June 17, 1994. Pp. 31402-31452.
5. Higgins, J. P. T., & Green, S. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley – Blackwell, New York. 2008.
6. The online Cochrane reports (Cochrane Collaboration). <http://www.cochrane.org>
7. Rothstein, H. R., Sutton, A. J., & Borenstein, M. *Publication Bias in Meta-Analysis*. Wiley & Sons, West Sussex, UK. 2005.
8. www.Meta-Analysis.com (This is the site from which this program was obtained. I used this program in writing this paper; it is a very useful system and easy to use.)
9. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effects and random-effects models for meta-analysis. *Res. Syn. Meth.*, 97 – 111. New York: John Wiley & Sons, Ltd.
10. Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, R. r. (2000). Empirical assessment of effect of publication bias on meta-analysis. *Biostat. Meth. J.* 320: 1574 – 1577.
11. Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. London: Academic Press.

Some Basic Points Concerning Meta-Analysis and Clinical Trials

About the Author:



Dr. Daryl S. Paulson, President and CEO, has extensive experience in skincare research designs, clinical trials, and biostatistics. He is the author of the standard texts of the industry -- [Topical Antimicrobial Testing and Evaluation](#), [Applied Statistical Designs for the Researcher](#), and [Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries](#) -- as well as the editor of the [Handbook of Topical Antimicrobials: Industrial Applications in Consumer Products and Pharmaceuticals \(Manufacturing Engineering and Materials Processing\)](#). Dr. Paulson has designed the procedures used at BioScience Laboratories, Inc., for evaluation of skin care and cosmetic products, as well as the statistical models used to assess the data. These include factorial designs, Analysis of Variance designs, regression analysis, exploratory data analysis, and integrative and statistical design in both parametric and nonparametric methods.

He is a member of the [American Society for Testing and Materials Subcommittee](#) that concerns itself with antiseptic-testing, the [American Society for Microbiology](#), the [American Organization of Analytical Chemists](#), the [Association for Practitioners in Infection Control](#), the [Parenteral Drug Association](#), the [American Society for Quality Control](#), [American Psychological Association](#), and the [American College of Forensic Examiners](#).

Dr. Paulson is the author of seven books and has published articles in several industry journals including, but not limited to [American Journal of Infection Control](#), [AORN Journal](#), [Applied and Environmental Microbiology](#), Dairy, Food & Environmental Sanitation, [Environmental Health](#), Food & Environmental Sanitation, [Food Protection Trends](#), [Food Quality](#), [Medical Device & Diagnostic Industry](#), [Pharmaceutical Engineering](#), [Journal of the American Dental Association](#), [Lab Manager Magazine](#), [Pharmaceutical & Cosmetic Quality](#), [Soaps & Cosmetics](#), and [Soaps/Cosmetics/Chemical Specialties](#).

Dr. Paulson welcomes your questions about how BioScience Laboratories can design a study specifically for marketing your product.

Daryl S. Paulson, Ph.D.
CEO and Founder
BioScience Laboratories, Inc.
Bozeman, Montana
dpaulson@biosciencelabs.com

**Some Basic Points Concerning
Meta-Analysis and Clinical Trials**

(406) 587-5735