Meta Analysis · 1
How to combine research studies

Chuck Chakrapani

Dealing with imperfect information
All marketing research studies are subject to sampling errors. After all, even if we use a reasonably large sample size of 1,000, our margin of error is ±3% or a spread of 6 percentage points. Consequently, if 2% of your customers are dissatisfied with the service provided by your firm in one study and this figure increases to 5% in the next study conducted 6 months later, it is not even clear whether your firm is really achieving a better performance. Even worse, if in the next survey the figure may go back where it was before. Obviously this raises questions:

- has your performance gone down (or up)?
- is nothing really happening?

While such questions are usually answered with the help of other information that may be available to the researcher or to management, it is clear that the data 'don't' speak for themselves'.

If we have unlimited money (and not enough intelligence to match) we can do enormously large studies at great expense. But that is not the real solution. Firstly, no matter how large a study is, there will always be sampling error, albeit small. Secondly, such resource outlay is seldom justified in marketing research where no meaningful marketing decisions can (or more importantly, should) be taken on the basis of a few percentage point difference in response to survey questions.

Consolidating what we know
What we really need may not be very large scale studies but methods which will enable us consolidate the knowledge we gathered over a period of time. This called 'meta-analysis'. Meta-analysis has several advantages.

- Meta-analysis may be able to offer definitive conclusions out a series of less than definitive studies.
- It saves resources by avoiding redundant studies-there is no need to carry out new studies if the main objective of the proposed study can be met through a re-analysis of the existing studies.
- Even when one-shot studies provide reliable information, there is no substitute for cumulative knowledge. Meta analysis provide cumulative knowledge.
- Meta-analysis can combine studies which may not be directly comparable. For example, in one study you may have used a five-point scale and in another study a seven-point scale.

These and other advantages of meta-analysis make it particularly attractive to those who constantly have to work under the constraints of limited resources.

Combined tests
The fundamental technical concern in any meta analysis relates to the procedures that can be used to combine different studies. Let us start with an example of a new marketing manager in a corporation confronted with four different studies carried out prior to her arrival. The results are given in Exhibit 1.

Exhibit 1
Results of 4 studies

Study	N	Avg. rating of corporaton	Ave. rating of competitor
1*	220	5.9	5.6
2**	250	3.6	3.9
3**	180	3.8	3.6
4*	150	6.2	5.7

* On a 7-point scale
** On a 5-point scale

From the Exhibit we note that Studies 2 and 3 used a five-point scale while studies 1 and 4 used a 7-point scale. In Studies 1 and 4 your corporation is rated better while in Study 2, the position appears to be reversed. In Study 3, the results are too close to be conclusive.

In short what we have is four studies with scales that are not directly comparable and results that are apparently inconclusive.

How do we combine the results? Fortunately, there a few simple tests which enable us to combine the results of different studies. We will discuss here three such tests.

1. Fisher Combined Test
Fisher Combined Test is given by
$c2 = - 2 \log_e p$
meaning

- obtain the one tailed probability (ie statistical significance) levels of each survey result;
- calculate the natural logarithm of these probabilities;
- multiply the result by - 2;
- add them together;
- The resultant number is a c2 value, distributed with degrees of freedom = 2 times the number of studies combined.

In our example, some studies favoured our corporation and other studies favoured the competition. How do we use the Fisher Combined Test to combine all four studies to get a better reading of what we know so far? (The steps involved are fairly simple and straightforward, but they require a knowledge of elementary statistics covering t and z statistics, degrees of freedom or df and c2 tests.)

In practical terms this is what we would do:
1. Calculate the t value (for differences) for each survey separately.

2. Find out the corresponding p (probability) value for each t. (A t value that is significant at the 95% level, for instance, will have a p value of .05.)

3. Consult the table of natural logarithms to compute - 2 loge p.

4. Add the values obtained in Step 4 to get - 2 loge p which is the c2 . This c2 value is distributed with df=8 (No. of surveys x 2).

These steps are laid out in Exhibit 2. When we consult the c2 table, the obtained value of 22.7 for 8 degrees of freedom is significant at the 99% level.

Exhibit 2
Fisher Combined Tests (Worked out example)

Study	N	Avg. rating of corporaton	Ave. rating of competitor	t*	p	- 2 $\log_e$ p
1	220	6.9	6.6	1.6	.055	5.80
2	250	3.6	3.9	-2.0	.975	0.06
3	180	3.8	3.6	1.6	.055	5.80
4	150	7.2	6.7	2.7	.004	11.04

x_2 = - 2 sum $\log_e$ p= 22.70
with 8 df

* one-tailed test

Although individual studies gave inconclusive results, when we combine all four studies we find that, overall, our corporation is rated more favourable compared to our competitor. We can consider this finding as a form of consolidation of what we have learnt so far.

Exhibit 3
Stouffer Combined Test

Study	N	Avg. rating of corporaton	Ave. rating of competitor	Z
1	220	6.9	6.6	1.6
2	250	3.6	3.9	-2.0
3	180	3.8	3.6	1.6
4	150	7.2	6.7	2.7

Z = 3.9
Zc = Z / N = 3.9 / 2 = 1.95

2. Stouffer Combined Test
Stouffer Combined Test is even easier to calculate. The formula is:
Zc = Z / N

where

Zc = z value for the combined test
N = Number of studies combined.

The procedure consists of the following steps.
1. Compute the z value for each survey.

52. Add them together.

3. Divide by the square root of the number of surveys.

The resultant value is Zc , whose level of significant can be computed by simply consulting the Z table.

In our case, Zc = 1.95. This is statistically significant at the 90% level.

(Note that the z values are identical to the t values in our example. This is because of the relatively large sample sizes we use in marketing research studies. The difference between the t and the z values can be large when we deal with small sample sizes i.e. sample sizes under 30 units.)

The significance of the Z value of the combined test can easily be assessed by referring to the standard Z distribution tables.

3. Winer Combined Test
The Winer Combined Test is yet another way of combining different studies. The Winer formula for combining is

$Zc = t / [df/(df-2)]$

Note that this formula is similar to the Stouffer formula with the following differences:

- We sum the t values instead of the Z values (although, as we discussed earlier, this makes very little difference for most marketing research survey results);
- The denominator consists of degrees of freedom for each sample (ie. sample size minus statistical freedom used) divided by (df - 2), which in our case should be close to 1 for each survey. Consequently, the resultant number will be very close to the square root of the number of surveys, the denominator used in the Stouffer Combined Test.

The calculations are shown in Exhibit 4.

Exhibit 4
Winer Combined Test

Study	N	Avg. rating of corporaton	Ave. rating of competitor	t	df	df - 2
1	150	7.2	6.7	1.6	149	148
2	250	3.6	3.9	-2.0	249	248
3	180	3.8	3.6	1.6	179	178
4	220	6.9	6.6	2.7	219	218

Sum t = 3.9
Zc = sum t / sum [df/(df-2) = 1.945

Which test to use?
In this article, we discussed three ways to combining different studies. Which test is better?
Let us compare the results obtained in the three tests:

Fisher	22.70	99%
Stouffer	1.95	99%
Winer	1.945	99%

(Note: Stouffer and Winer tests fall slightly short of 99%, but close enough to be considered significant at the 99% level)

First, let us compare the Stouffer and the Winer tests. The results are identical. This is no accident. Because of the (relatively) large samples used in marketing research, t and z values converge and (df/df-2) gets very close to 1. Because of this, we are likely to get practically the same results whether we use the Stouffer or the Winer test. However, because, Stouffer is somewhat easier calculate than Winer, so on most situations one may prefer Winer to Stouffer.

The Fisher Combined Test is considered to be 'asymptotically efficient'. Since it deals with the p values rather than with specific statistical values (such as t or Z), you can use it combine studies which have used different statistical procedures.

In general, the differences among these different tests are not too large. In most situations, these testa can be sued interchangeably. No matter which test you use, you should get similar results.

The list of references that follow contains more information on combined tests discussed here. You may note from the references that, although meta-analysis has been gaining in popularity only recently, the statistical concepts on which it is based have been around for a long time. No new major concepts are involved in combining different studies.

References for Combined Tests (Meta-analysis)

Fisher Combined Test
Fisher (1932) Statistical Methods for Research Workers (4th ed.) London: Oliver & Boyd.
Fisher (1948) Combining independent tests of significance. American Statistician. 2(5):30.

Winer Combined Test
Winer, B.J. (1971) Statistical Principles in Experimental Design (2nd ed.). New York: McGraw-Hill

Stouffer Combined Test
Stouffer S.A., Suchman, E.A., De Vinney, L.C., Star, S.A. & Williams, R.M. Jr. (1949) The American Soldier: Adjustment During Army Life, Vol. 1. Princeton, NJ: Princeton Univ. Press.

Mosteller, F.M. & Bush, R.R. (1954) Selected Quantitative Techniques. in Lindzey, G. (ed.) Handbook of Social Psychology, Vol.1. Cambridge, MA: Addison-Wesley.

Rosenthal, R. (1978) Combining the results of independent studies. Psychological Bulletin. 65:185-193.

Chuck Chakrapani is President of Standard Research Systems Inc. He is the author of several books and is the Editor-in-Chief of the Canadian Journal of Marketing Research.