Florida Journal of Educational Research
Fall 2000 Vol. 40, No. 2, 76 – 99.

Risky Predictions and Damn Strange Coincidences:

An Initial Consideration of Meehl's Index of Corroboration

Kristine Y. Hogarty

University of South Florida

The explication and empirical testing of theories are critical components of research in any field. Despite the long history of science, the extent to which theories are supported or contradicted by the results of empirical research remains ill defined. Meehl (1997) has proposed an index of corroboration (C_i) that may provide a standardized means of expressing the extent to which empirical research supports or contradicts a theory. The index is the product of a theory's precision of prediction and the extent to which observed data are close to those predictions. Large values of C_i are expected from strong theories making tight, accurate predictions. Small values should result from (a) weak theories making weak predictions (regardless of their accuracy), or (b) strong theories that are not accurate.

Simulation methods were employed to evaluate the sampling behavior of C_i. Factors in the research design included the precision of prediction, degree of congruence between known population parameters and the theoretical prediction, sample size, psychometric reliability and the influence of a confounding variable. The results suggest that precision of prediction is far more influential in the value of C_i than is the accuracy of prediction. As anticipated, less reliable measures yielded smaller values of C_i. An uncontrolled extraneous variable resulted in biased C_i values, but the direction of bias could not be anticipated. Surprisingly, sample size evidenced a negligible influence on the average value of C_i, although sampling error was reduced with larger samples.

The explication and empirical testing of theories are important components of research in any field. Kerlinger (1964) suggested that these components are fundamental distinctions between science and common sense. "While the man [sic] in the streets uses theories and concepts, he ordinarily does so in a loose fashion...The scientist, on the other hand, systematically builds his theoretical structures, tests them for internal consistency, and subjects aspects of them to empirical test" (p. 4).

However, despite the long history of science, tools for explicating the extent to which theories are supported or contradicted by the results of empirical research remain ill defined. Often such support or contradiction is reduced to the "reject" or "fail to reject" decisions resulting from tests of null hypotheses that are derived from aspects of theory. That is, a theory is "supported" by empirical evidence if null hypotheses are rejected, when the theory suggests they should be rejected. Conversely, a theory is contradicted (and may be considered "refuted," cf. Popper, 1959) if such theoretically derived null hypotheses are not rejected. The limitations of null hypothesis testing are well known (viz., Harlow, Mulaik, & Steiger, 1997), but its use in the testing of theories presents unique conceptual challenges and interpretational dangers.

In recent years, such an overly simplified approach to theory testing has been challenged on logical grounds (Meehl, 1978, 1990, 1997; Serlin & Lapsley, 1985). The essential aspects of these logical arguments are twofold. First, theories differ in the extent to which they provide precise predictions about observations. For example, a prediction that middle schools boys and middle school girls will have different means on some variable is a relatively weak prediction. A prediction that the mean of girls will be greater than that of boys is somewhat stronger, while a prediction that the means will differ by some value between nine and 15 points is stronger yet, and a prediction that the means will differ by exactly 12 points is even more precise. The precision of predictions derived from theories is proportional to the strength of support that may be provided by empirical evidence congruent with the prediction. That is, a precise prediction that is supported by data provides more logical evidence in support of the theory than does a weak prediction supported by data.

This relationship between the precision of prediction and the strength of logical support is rooted in the relative rarity of the data, absent the theory. That is, without the theory, would we expect to see such data anyway? The extent to which we would *not* expect to see such data is what Salmon (1984) refers to as a "damn strange coincidence," and the extent to which a theory predicts such otherwise rare data is a "risky prediction" (Meehl, 1978).

Further, the movement from theory into an empirical test necessitates the incorporation of many logical components besides the theory itself. It is the incorporation of these elements which distinguishes theory testing from a test of some statistical hypothesis H_0. Meehl (1997) presents these components as elements of an equation:

$$\left(T \bullet A_t \bullet C_p \bullet A_i \bullet C_n\right) \rightarrow \left(O_1 \supset O_2\right)$$

where T = the theory being "tested,"

A_x = Auxiliary theories relied upon during the conduct of the research,

C_p = *Ceteris paribus* (all other things being equal),

A_i = Instrumental theories related to measures and controls employed,

C_n = Realized particulars, the extent to which the research was actually

conducted as we think it was, and

$(O_1 \supset O_2)$= the material conditional "if you observe O_1, you will observe O_2."

That which is subject to empirical test is not the theory alone, but the amalgam of these elements. Data which appear to contradict a "theory" may arise because of errors anywhere in this combination of elements (e.g., the theory may be correct but the groups we thought were equivalent were actually systematically different from each other on an important, confounding variable).

Auxiliary theories (A_x) lie at the periphery of the theory being tested and are somewhat distinct from the "hard core" concepts or postulates of the theory under investigation. Although central portions of a particular theory may not be rigorously defined, there will likely exist key critical components as well as non-central elements. These tangential components (although not central to the theory being explored) are still, in fact, a part of the theory.

For example, in an investigation of the relationship between nutrition and anxiety in which anxiety is measured using responses to Likert-type items written in English, the use of participants whose primary language is not English necessitates an auxiliary theory that the anxiety instrument retains its validity in such a population. If data obtained from such research fail to support theoretical predictions, the failure may be attributable to the core theory being incorrect or simply that the auxiliary theory did not hold.

The concept of verisimilitude (truth-likeness) is closely related to this core-peripheral distinction. Meehl (1990) suggests that a theory that is false in its core postulates has lower verisimilitude than one that, while correct in its core concepts, is incorrect in several of its peripheral ones. As even the best theories are likely to be approximations of the true state of reality, verisimilitude then, refers to the relationship between the theory and the real world.

Ceteris paribus does not mean that all factors not mentioned are equal for all participants, rather that there are no systematic factors left unmentioned. This clause amounts to a very strong and highly improbable negative assertion that "nothing else is at work except factors that are totally random and therefore subject to being dealt with by our statistical methods" (Meehl, 1990, p. 111).

The instrumental auxiliary theories (A_I) are related to measures and controls employed by the researcher. These are distinguished from A_x in that they do not contain any psychological constructs. Thus, if anxiety is measured by changes in galvanic skin response rather than by a Likert instrument, the auxiliary theory at work is within A_I rather than A_x.

The realized particulars (C_n) represent the extent to which the research was actually conducted as we think it was. This element of the amalgam represents treatment integrity. For example, if we plan to manipulate participant nutritional status to examine its relationship with anxiety, but the participants do not adhere to their dietary "treatment," then the variable actually applied in the research is not what we think it is. Data that contradict our theory may arise because of this perturbation in C_n.

Meehl's Index of Corroboration

Meehl (1997) has proposed an index of corroboration (C_i) that may provide a standardized means of expressing the extent to which empirical research supports or contradicts a theory:

$$C_i = (Cl)(In)$$

where *Cl* = the "closeness" of the data to the theoretical prediction, and

In = the "intolerance" of the theory (e.g., a standardized precision of prediction).

These terms are further explicated as follows:

$$Cl = 1 - (D/S)$$

where *D* = deviation of observed data from the tolerance interval of the theory

S = Speilraum (the range of data values that are expected whether or not the theory is true)

$$In = 1 - (I/S)$$

where I = the interval tolerated by the theory (e.g., the raw precision of prediction).

The index is thus the product of a theory's precision of prediction and the extent to which observed data are close to those predictions. Large values of C_i are expected from strong theories making tight predictions in which data are very similar to predicted values. Small values should result from (a) weak theories making weak predictions (regardless of the congruence of the data with those predictions), or (b) strong theories making tight predictions in which the data are not congruent with the predictions.

In order to elucidate the expected behavior of Meehl's corroboration index, an earlier example is extended. Recall that large values of C_i should result from strong theories making tight predictions in which data are very similar to predicted values. Returning to our earlier example, let's suppose a researcher has made a prediction that middle school girls will score higher than middle school boys on a given measure of self-esteem. This prediction is somewhat stronger than a prediction that middle schools boys and middle school girls will have different means on this measure, because a direction of difference is predicted. However, the prediction is less precise than a prediction that the means will differ by some value between 5 points and 9 points, with the girls presenting a higher mean than the boys. Further, suppose that the plausible values of mean difference, whether or not the theory is true, range from -10 to +10. The Spielraum (S) is thus 20.

In this example, the simple directional prediction of higher means for girls suggests a tolerance interval of 10 points (any mean difference greater than zero is consistent with this "flabby" prediction) and an intolerance (In) of 1 - 10/20 or 0.50. If the sample mean for girls is found to be 6.0 points higher than that of boys, the data do not deviate from the prediction (Cl = 1.0) and Meehl's C_i = (Cl)(In) = (1.0)(.50) = .50. If the prediction was not simply "girls greater than boys," but "girls between 5 and 9 points greater than boys," then the tolerance interval is 4 points and In = 1 – 4/20 or .80. The same observed data (a difference in means of 6.0 points) are also consistent with this prediction, but C_i = (1.0)(.80) = .80. The latter theory receives more corroboration from the data because it made a riskier prediction that was consistent with the observations.

Suppose the observed data evidence a 2.0 point difference in which the middle school boys scored higher than the middle school girls. Such data are not consistent with the predictions of either theory. For the theory providing a directional prediction only, the data deviate (D) from the lower bound of the tolerance interval by 2.0 points and Cl = 1 – D/S = .90. These data provide a corroboration index value of (Cl)(In) = (.90)(.50) = .45. For the riskier prediction of a difference between 5 and 9 points (favoring girls) the data deviate by 7 points and Cl = 1 – D/S = .65. For this theory, the data provide a corroboration index value of (Cl)(In) = (.65)(.80) = .52. Although the observed data deviate to a greater extent from the prediction of the latter theory, the corroboration is still greater because the prediction was more precise. Figure 1 presents the values of Meehl's C_i that would result from any observed mean difference within the range specified by the Spielraum.

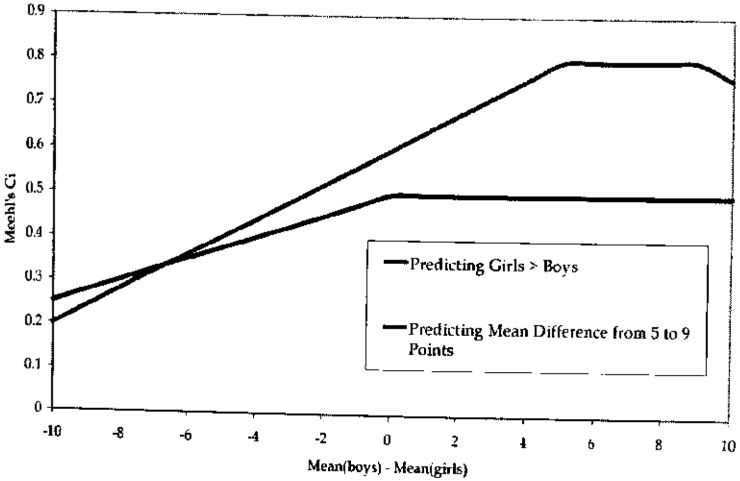

Figure 1. *Meehl's C_i by Observed Mean Difference*

Note that the theory making a more precise (riskier) prediction receives more corroboration than the flabbier theory unless the observed data yield a mean difference of more than six points in a direction opposite that of the prediction. Further, note that the intolerance of the theory (In) presents an upper limit for Meehl's C_i.

Purpose of the Study

Meehl (1997) has presented a logically sound index of corroboration to summarize the extent to which empirical tests of theories provide support or contradiction to those theories. However, the numerical properties of this index have not been investigated. The research to be reported represents an initial venture into the exploration of this index and its behavior in the testing of theories in the social sciences.

Method

The behavior of Meehl's C_i was evaluated using Monte Carlo methods. A series of simulations were conducted that related theoretical predictions to empirical results. The use of simulation methods allows the control and manipulation of research design facets and the incorporation of sampling error into the analyses. The study was designed in the context of a simple theory, the core of which predicts a difference in means between two groups.

Eight factors were manipulated in these simulations: factors related to the nature of the theory being tested, the degree of correspondence of the theory to the actual populations simulated and research design factors. First, three factors related to the theory being tested were included. The predicted mean difference between groups was examined at five levels (0.00, 0.25, 0.50, 1.00 and 2.00), the raw tolerance interval of the theory was examined at four levels (0.25, 0.50, 1.00 and 2.00), and the Spielraum was examined at three levels (4, 8 and 16). These values of raw tolerance and Spielraum yield intolerance (*In*) values ranging from 0.50 (the value of intolerance for a simple directional prediction of effects) to 0.98 (reflecting a tight, risky prediction).

Second, two factors related to the true populations simulated were manipulated. The population difference in means was examined at five levels (0.00, 0.25, 0.50, 1.00 and 2.00), and variance ratios between the two populations were manipulated at four levels (ratios of 1:1, 2:1, 4:1 and 8:1). These population mean differences, crossed with the theory's predictions provided conditions ranging from those in which the theory's prediction exactly represented the true populations (perfect verisimilitude), to those in which the theory deviated from the true population conditions by effect sizes as large as two standard deviations.

The relationship between theory precision and theory verisimilitude (truth-likeness) was framed in a variety of research contexts, representing the other elements of the amalgam that is tested in research:

$$\left(T \bullet A_t \bullet C_p \bullet A_i \bullet C_n\right) \rightarrow \left(O_1 \supset O_2\right)$$

Specifically, three factors related to the design of empirical research were included in the simulations. Sample size was examined at five levels (5, 10, 50, 100 and 500 observations per group) and the reliability of the dependent variable was examined at five levels (r_{xx} = .40, .60, .80, .90 and 1.00). Finally, the confounding effect of an extraneous variable was examined at five levels.

To manipulate the reliability of the dependent variable, measurement error was simulated in the data (following the procedures used by Maxwell, Delaney, & Dill, 1984; and by Jaccard & Wan, 1995), by generating two normally distributed random variables to produce an observation (one to represent the "true scores" on the dependent variable, and one to represent measurement error). Fallible, observed scores on the dependent variable were calculated (under classical measurement theory) as the sum of the true and error components. The reliabilities of the scores were controlled by adjusting the error variance relative to the true score variance

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

where σ_T^2 and σ_E^2 are the true and error variance, respectively, and $\rho_{xx'}$ is the reliability. In this study, dependent variable reliabilities of 0.40, 0.60, 0.80, 90, and 1.00 were examined.

The influence of a confounding variable was included in the design to examine the effects of violations of *ceteris paribus* on the values of C_i. The data for the simulations were generated from the linear model

$$X_{ijk} = \mu + \alpha_j + \beta_k + \varepsilon_{ijk}$$

$$\begin{aligned} &\textit{where } X_{ijk} = \textit{observed value} \\ &\quad \mu = \textit{grand mean} \\ &\quad \alpha_j = \textit{population effect} \\ &\quad \beta_k = \textit{effect of extraneous factor, and} \\ &\quad \varepsilon_{ijk} = \textit{random error} \end{aligned}$$

The value of β_k was manipulated to produce effects of a confounding factor in the research design. Specifically, β_k was set to both positive and negative values equal in magnitude to α_j and equal to half the value of α_j.

Finally, conditions with $\beta_k = 0$ were included to represent controlled experiments that evidenced no confounding factors.

A simulation of empirical research was conducted and the resulting evidence (i.e., C_i), evaluated while manipulating the influence of confounding variables (C_p), and the reliability of instrumentation (A_x). These design facets were modeled in the simulations, respectively, as variations in pre-existing group differences and random errors of measurement in the criterion variable. For each condition examined, 50,000 experiments were simulated. The data resulting from each experiment were pooled and the average value of C_i was evaluated in the context of the central design factors.

Results

Initially, the results were examined with regard to three design factors in the study: sample size, intolerance and verisimilitude. This is an important consideration given that the interaction of these integral components of Meehl's corroboration index provides insight regarding the structure underlying the theory. The relationships between these factors are illustrated in a series of figures. Figure 2 presents the relationship between mean C_i value and both verisimilitude and intolerance averaged across all sample sizes for a Spielraum of 4.

This figure clearly demonstrates that the level of intolerance is a more salient influence than the degree of verisimilitude. As the level of intolerance is increased we see substantially more evidence of corroboration. The relationship between mean C_i, intolerance (In) and verisimilitude (Cl), [C_i =(Cl)(In)], is such that low levels of intolerance result in less evidence of corroboration. This relationship, however, appears to be moderated by the accuracy of the prediction. As the theoretical predictions deviated from the true population mean difference, less change in mean C_i was evidenced across intolerance levels.

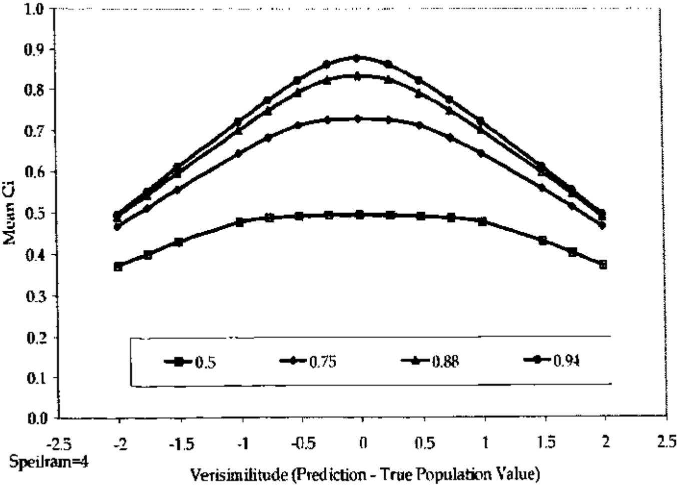

Figure 2. *Mean Value of Meehl's C_i by Intolerance and Verisimilitude.*

For those predictions that were relatively accurate (i.e., close to the true population mean difference) a much greater difference in mean C_i was witnessed than when the deviation from theoretical prediction was more pronounced. For example, when theoretical predictions were exactly correct (verisimilitude=0), the mean C_i was estimated to be .87 for an intolerance=.94, but dropped to .49 for an intolerance=.5. However when verisimilitude was 2 (a 2-point difference between the prediction and the true population mean difference), the mean C_i was estimated to be .50 and .37 when intolerance=.94 and .50, respectively. Similar patterns were witnessed at each level of intolerance with the 0.50 level of intolerance evidencing the least amount of change across varied levels of verisimilitude. These results suggest that intolerance is most influential when the theory is close to the truth.

Sample Size and Verisimilitude

The relationship between Meehl's C_i and both verisimilitude and sample size is illustrated in Figure 3. Examination of this figure reveals the negligible influence that sample size has on mean C_i. As sample size was increased, modest increases in mean C_i were seen when the prediction was very close to truth (verisimilitude within the range of -1 to +1). The estimated change in mean C_i in these cases was approximately .08 (e.g., a change from .68, N=5 to .76, N=500, verisimilitude=0). Substantially less change was evidenced with greater departure from truth (i.e., verisimilitude >± 1).

The relationship between the standard deviation of Meehl's C_i by verisimilitude and sample size is presented in Figure 4. Here we see slight increases in standard deviation as the theory moves away from truth.

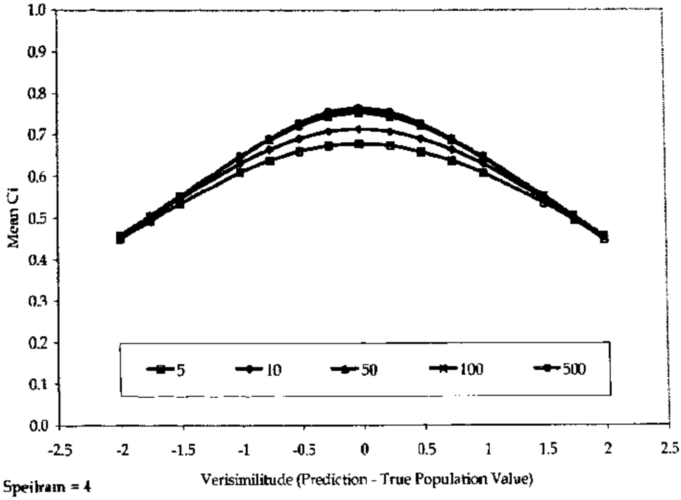

Figure 3. *Meehl's C_i by Verisimilitude and Sample Size.*

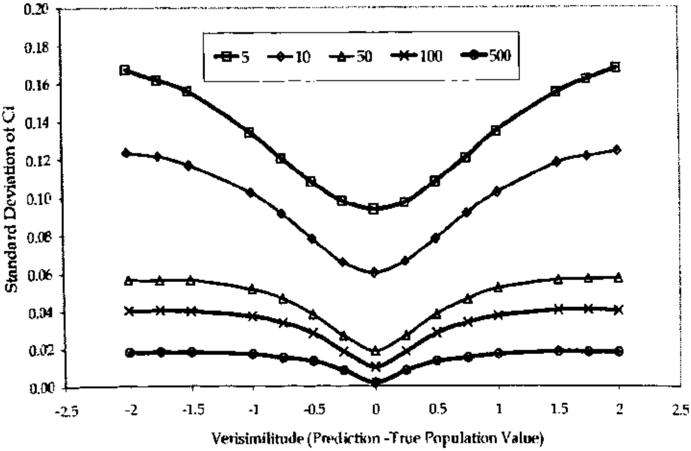

Figure 4. *Standard Deviation of Meehl's C_i by Verisimilitude and Sample Size.*

Further, the magnitude of the increase is greater with smaller samples (relatively flat curves with n = 500, with steeper curves as sample size decreases). However, there was substantially less variability associated with the sample C_i in the larger samples than with smaller samples, as indicated by the smaller standard deviations associated with the C_i of the larger groups. That is, sampling error is reduced with larger samples, resulting in less inter-sample variability in the index. However, the average value of the corroboration index was not appreciably influenced by the size of the sample.

Verisimilitude and Reliability

The relationship between Meehl's C_i and both verisimilitude and reliability of measurement is illustrated in Figure 5. Examination of this figure reveals the limited impact that the reliability of the dependent variable has on mean C_i. This is similar to the effect that was seen when the relationship between sample size and verisimilitude was examined. As reliability was increased, only slight increases in mean C_i were witnessed when the prediction was very close to truth (verisimilitude within the range of -1 to +1). With greater departures from truth, there was essentially no change evidenced in mean C_i across the various levels of reliability.

The relationship of reliability to the standard deviation of C_i (Figure 6) is also similar to that observed with sample size. That is, the use of reliable measures of the dependent variable resulted in greater consistency of the corroboration index across samples.

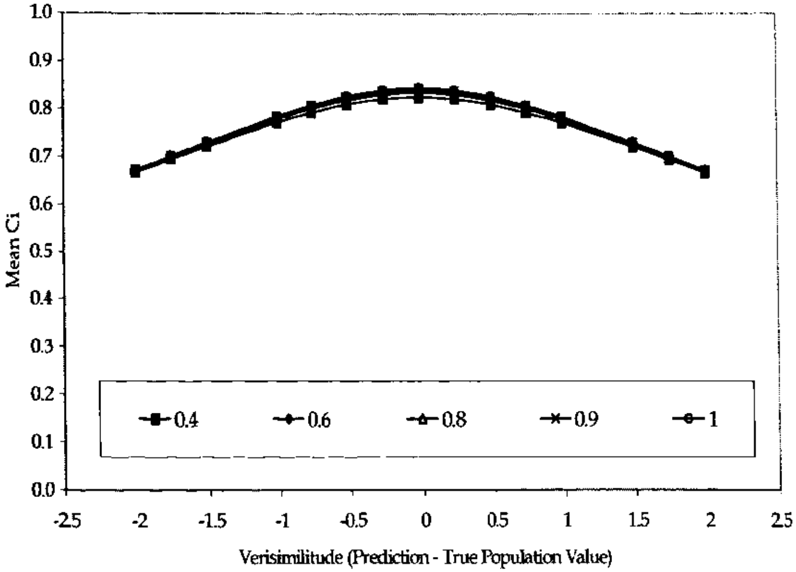

Figure 5. *Mean Value of Meehl's C_i by Verisimilitude and Reliability.*

Influence of Extraneous Variable

Figure 7 illustrates the effect of an extraneous variable on the mean value of C_i. These data were obtained from conditions in which the true population mean difference was 1.0, and a variety of theoretical predictions of the mean difference were tested (theoretical predictions are plotted on the abscissa).

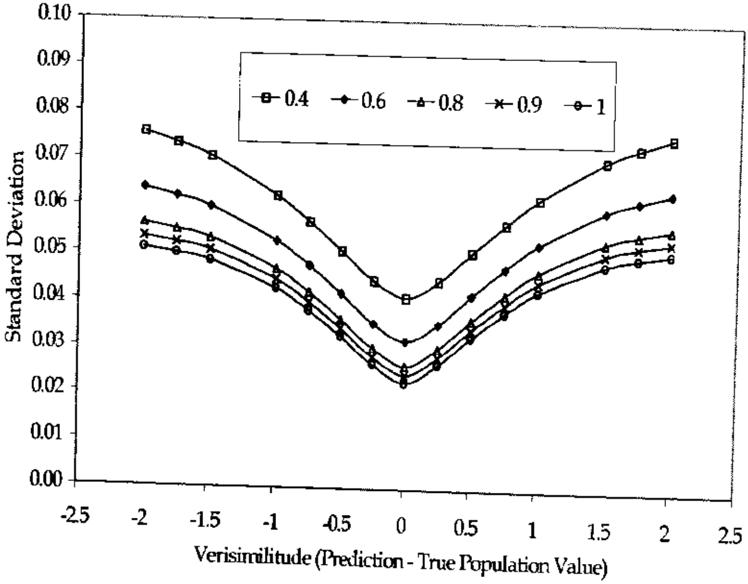

Figure 6. *Standard Deviation of Meehl's C_i by Verisimilitude and Reliability.*

The dashed line in the figure provides the mean value of C_i when extraneous variables are controlled. In this condition, C_i reaches its maximum value when the theoretical prediction is 1.0, coinciding with the true population mean difference. At this value of predicted mean difference, the influence of an uncontrolled extraneous variable resulted in a reduced value of C_i, regardless of the direction of the extraneous variable's effect.

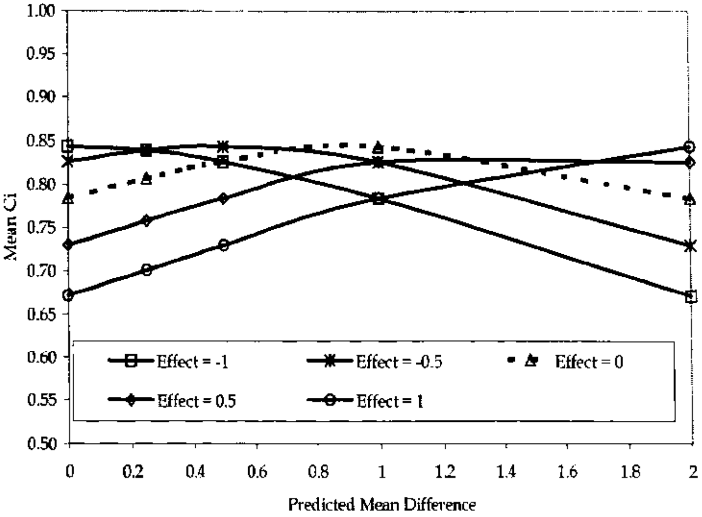

Figure 7. *Mean Value of C_i by Theory Prediction and Effect of Extraneous Variable. True Mean Difference = 1.0.*

However, when the prediction was not accurate, the effect of the uncontrolled extraneous variable either reduced or inflated the mean C_i value. For example, with a theoretical prediction of a 2.0 point difference in means (a prediction that is 1.0 point too high), an uncontrolled extraneous variable that increases the difference between sample means (an effect of +0.5 or +1.0) yielded a higher value of the corroboration index than that obtained when the extraneous variable was controlled. Conversely, an uncontrolled extraneous variable that reduced the difference in means resulted in data that were further removed from the predicted difference, yielding a negatively biased C_i value. Thus, uncontrolled extraneous variables result in biased C_i values, but the direction of bias cannot be determined unless the theoretical prediction is accurate. Such conditions consistently result in negative bias in C_i.

Conclusions / Recommendations

This study was designed to investigate the relationship between theoretical predictions and empirical results through an initial consideration of Meehl's index of corroboration. As the numerical properties of this index have not been previously investigated, our efforts were aimed toward illuminating the relationship between the closeness of the observed data or verisimilitude and the precision of prediction. An important limitation of this research is that our initial investigations have considered these relationships for only the most basic of predictions, that is, predictions about population mean differences. Under these very limited circumstances, the mean index of corroboration was seemingly unaffected by sample size, and notably more influenced by the level of verisimilitude and the level of intolerance specified by the theory. In addition, the reliability of the dependent measure was shown to have but a slight influence on the mean C_i, and only when predictions were very close to truth. Although of little impact with these basic predictions, measurement error would be expected to have a more noticeable influence in the context of more complex theoretical predictions (contexts involving partial correlations or multivariate analyses in which measurement error results in statistically biased estimates). The influence of a confounding variable was seen to yield either positive or negative biases in C_i depending upon the direction of the variable's effect and the relative inaccuracy of the theoretical prediction.

Although sample size and measurement reliability were not important determinants of mean C_i, both factors were related to the variability of this statistic, with larger samples and more reliable measures providing greater stability across samples. Although such sampling variability is important, one would anticipate that the degree of support for a theoretical prediction that was tested with a large sample should be greater than that provided by a small sample. Future work should be aimed at incorporating a sample size component into an index such as Ci.

Additionally, the influence of the other factors in the amalgam, and their relationship to the corroboration index remains to be investigated. These other elements need to be explored and tested under different conditions. For example, further work could be directed toward investigating the effects of such elements as threats to internal validity (A_x) and operational integrity (C_n).

Lastly, the relationship between intolerance and verisimilitude will need to be examined in a multitude of research contexts. A theory's merit is a matter of degree rather than a yes or no question, as it is treated in null hypothesis testing (Meehl, 1990). A natural extension of this work would be the examination of these relationships when making more complex predictions from theories. For example, multivariate extensions of C_i can be investigated in the context of path analyses or structural equation modeling. An extension of the components of C_i to multivariable problems is worthy of investigation. For example,

$$Intolerance = 1 - \prod_{j=1}^{J} \frac{I_j}{S_j}$$

$$Closeness = \left[\prod_{j=1}^{J} \left(1 - \frac{D_j}{S_j} \right) \right]^{\frac{1}{J}}$$

where j indexes the set of relationships being tested (i.e., I_j and S_j are the tolerance interval and Spielraum for variable j and D_j is the distance between the theoretical value and the observed value).

Such a multivariable index may be superior to traditional indices of "fit" used in applications such as structural equation modeling. Rather than focusing on the fit between a structural model and an observed covariance matrix (which may be equally well or better fit by a large number of structural models), this index could be used to represent fit between predicted and observed structural parameters.

Although this research presents only the initial consideration of the behavior of Meehl's C_i, our results support the utility of the index. Meehl's C_i has applications for the planning of empirical studies as well as for the interpretation of research results. Its use should serve to move the arguments surrounding theory testing away from the testing of null hypotheses into a consideration of the complexity of the research context, the degree of "risk" entailed by the theory's predictions, and the extent to which the obtained data (absent the theory) represent a "damn strange coincidence." Finally, in a similar manner, the index may be useful in moving meta-analyses of research results beyond the stage of simple statistical aggregation of empirical results into a more appropriate (albeit complex) context.

References

Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348-357.

Kerlinger, F. N. (1964). *Foundations of behavioral research.* New York: Holt, Rinehart, and Winston.

Maxwell, S. E., Delaney, H. D. & Dill, C. A. (1984). Another look at ANCOVA versus blocking. *Psychological Bulletin, 95,* 136-147.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806-834.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1,* 108-141.

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (Eds.). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Basic.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, N.J.: Princeton University Press.

Serlin, R. C. & Lapsley, D. K. (1985). Rationality in psychological research: The good enough principal. *American Psychologist, 40,* 73-83.