Poster Presented at the 102nd Annual American Psychological Convention; August, 1994; Los Angeles

Validation of the Schizotypal Taxon Using Graphical Taxometric Techniques

Viktor Brenner and Michael L. Raulin
State University of New York at Buffalo

This study was designed to cross-validate both the presence of a hidden schizotypal taxon in a large college student population as well as a new taxon search procedure, MAXSLOPE. MAXSLOPE promises to be more user-friendly than earlier techniques because it uses a graphical approach that is easier to visualize. Previous research found a HITMAX (the point where two underlying distributions cross) for three measures of schizotypal signs—Magical Ideation, Perceptual Aberration, and Cognitive Slippage (Lowrie & Raulin, 1990). Base rate estimates from two studies hover around 10% (Lenzenweger & Korfine, 1992; Lowrie & Raulin, 1990), consistent with other research (Meehl, 1990). MAXSLOPE was used on data from 2591 undergraduates. We made a slight modification to make MAXSLOPE appropriate for integer data. When the data were analyzed without separating by sex, a HITMAX was found on the Magical Ideation Scale, but could not be found for the Perceptual Aberration or Cognitive Slippage Scales. The base rate of schizotypy was estimated to be 9.2%. However, when separated by sex, no HITMAX for males could be found on Magical Ideation. Females showed a HITMAX, but the estimate of base rate was 17%. Potential problems with these estimates are discussed, and hypotheses are ventured for the pattern of results found. The partial validation of previous results is suggestive and encouraging, but not conclusive.

Schizophrenia is probably the most studied disorder in psychopathology. Yet despite over 100 years of work, relatively little is known about the etiology of schizophrenia. In the early 60s, Paul Meehl proposed a diathesis-stress model of schizophrenia—one of the most influential models to date. This model postulated that all schizophrenics are born with a schizophrenic gene (schizotaxia), which leads to an aberrant personality organization (schizotypy). Given unfortunate environmental stressors, people with this personality organization develop schizophrenia. It is indicative of how enigmatic schizophrenia is that this model could be presented nearly 30 years later (Meehl, 1990), with greater detail, but basically unchanged.

Meehl's model has stimulated considerable research on schizotypy. However, many questions about the etiological role of schizotypy in the development of schizophrenia remain unanswered to this day. The first step towards validating the hypothesized causal relationship is to define and measure the schizotypal personality organization. Meehl (1964) proposed a checklist of symptoms that has served as the blueprint for such work, and the Chapmans and their students have developed several self-report measures of these schizotypal signs. Among these are scales for Physical Anhedonia (Chapman, Chapman, & Raulin, 1976), Perceptual Aberration (Chapman, Chapman, & Raulin, 1978), Magical Ideation (Eckblad & Chapman, 1983) and Cognitive Slippage (Miers & Raulin, 1985), although there are many others. Research (Chapman, Chapman, Raulin, & Edell, 1978; Propper, Raulin, Lowrie, Trigoboff, Henderson, & Watson, 1987) suggests that the Physical Anhedonia Scale identifies individuals who show a lack of emotional involvement. The Magical Ideation, Cognitive Slippage, and Perceptual Aberration Scales identify individuals characterized by sub-psychotic peculiarities of thinking. A long-term follow-up study has just been completed. The initial report (Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994) suggests that anhedonic subjects are not at risk, but that scales that identify sub-psychotic symptoms do identify subjects at heightened risk for future psychosis.

One of the features of schizotypy that is of interest to schizophrenia researchers is its epidemiology. It is generally accepted that not everyone who is genetically at risk for schizophrenia develops it. As a result, it is difficulty to get an estimate of how many people carry the genetic risk factor(s). Relative rates of genetic risk and disorder suggest differing transmission mechanisms (polygenetic, single-gene recessive, etc.), and so these numbers are of interest to genetic researchers. Meehl's theory predicts that about 10% of the population should be schizotypal (Meehl, 1990), and 10% of schizotypes (1% overall) should develop schizophrenia. However, schizotypy is subtle and cannot at present be accurately assessed by the interview methods usually used by epidemiologists. Our measures are at best only **moderately valid indicators** of the schizotypal personality organization. We have no way of knowing which of the identified cases are hits and which are false positives. We are left wanting to know the relative epidemiology of schizotypy and schizophrenia, but no direct way to get that information. Faced with this problem, methods have been developed to estimate population parameters indirectly. These mathematical procedures are collectively known as taxon search procedures.

Taxon Search Procedures

Lacking a perfectly valid indicator of schizotypy, we are forced to develop methods to estimate population parameters such as prevalence. The nomenclature for the procedures used to estimate parameters is unfamiliar to most psychologists, making it difficult to understand for someone not already familiar with the techniques. Therefore, we will define each term as we introduce it.

Taxon search procedures assume that an **observed distribution** on a moderately valid indicator variable of the taxon is a mixture of two **underlying distributions.** If the underlying distributions could be seen, a graph of them might look like Figure 1. These two underlying distributions represent separate populations or **taxonomic categories** (taxa, for short). According to Meehl's model of schizophrenia, the general population is made up of two taxa—schizotypal and non-schizotypal—with the estimated base rates shown in Figure 1.

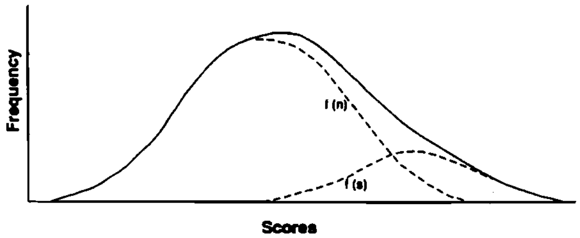

Figure 1 The hypothetical situation of the distribution of a moderately valid indicator of an underlying taxonomy (two categories with base rates of .9 and .1).

A **moderately valid indicator** of a taxonomy is a measure that shows a mean difference between the taxa, but with considerable overlap of the distributions. This contrasts with a perfectly valid indicator, where there is no taxon overlap. If a perfect measure existed, classification of subjects into the taxonomic categories is trivial, and there would be no need to estimate the parameters of the taxa. However, perfectly valid indicators seldom exist.

Taxon search procedures are methods that use the mathematical properties of moderately valid indicators to estimate the parameters of the taxa. Central to all taxon search procedures is the identification of the point at which the two underlying distributions cross, called the HITMAX. **HITMAX** is the decision criterion point that maximizes the number of correctly classified subjects (See Appendix A for an explanation of why this is the case).

Each taxon search procedure uses mathematical properties to estimate the HITMAX. One of the first taxon search procedure developed, as well as most widely used, is the **maximum covarariance procedure (MAXCOV**, for short; Meehl, 1973). However, difficulties in the implementation of this procedure may have limited its application to date. First, this procedure needs three moderately valid indicators of the same underlying taxon in order to determine the HITMAX. These indicators are assumed to be uncorrelated within taxa, although the procedures are robust to this assumption (Grove & Meehl, 1993). It is not always possible to find three such moderately valid indicators of a taxonomy. Second, in order to use MAXCOV, it is necessary to perform covariance analysis repeatedly on small subgroups of subjects. One scale is chosen as a base scale. The scale is subdivided into multiple overlapping units by a process know as taking SLIDING CUTS (see Appendix A for more details). The covariance between the two other indicators is calculated for those subjects who score in a given cut on the base scale. By finding the peak of covariance on the other scales, the HITMAX on the base scale is identified. Although straight forward, no existing data analysis package performs these computations easily. Therefore, one is forced to do tedious analyses or write software to automate the process. A final problem is that the output of MAXCOV is not always easily interpretable. For example, the output cannot be mapped onto a common metric such as an *F*-ratio. As a result of these problems, MAXCOV is probably underutilized.

Recognizing the potential inaccessibility of MAXCOV, Meehl and others have continued to refine taxon search procedures. At present, more than 40 such methods have been developed, requiring anywhere from one to seven or more indicators of taxonicity (Meehl & Yonce, 1989). One of the newest methods is MAXSLOPE. While still being tested, MAXSLOPE shows promise as a taxon search procedure because it yields more easily interpretable results while only requiring two moderately valid indicators. The key difference between MAXCOV and MAXSLOPE is that the latter is graphical in nature. Its output potentially offers more easily readable results. However, MAXSLOPE has not yet been subjected to extensive testing. One purpose of this investigation will be to use MAXSLOPE to cross validate parameter estimates of the schizotypal taxon arrived at by using MAXCOV, thus establishing convergent validity.

Schizotypy research is one of the few areas where taxon search procedures have been used. The existence of a schizotypal taxonomy (i.e., a taxonomic class of people at genetic risk for schizophrenia) is a subject of debate. However, finding a HITMAX in a distribution of indicators is suggestive of taxonicity, offering some degree of validation for the theory. At least two studies have identified HITMAX cuts using schizotypy scales (Lenzenweger & Korfine, 1992; Lowrie & Raulin, 1990). In both cases, base rate estimates hovered around 10%, a figure consistent with base rate estimates from genetic studies (Meehl, 1990).

The current study attempted to cross validate both a new taxon search technique and the current estimates of the size of the schizotypic taxon. If a HITMAX can be found and the parameter estimates of the size of the taxa are similar to the results of Lowrie and Raulin (1990) and Lenzenweger and Korfine (1992), it would strengthen the argument for the existence of a schizotypal taxon. A successful search would show that MAXSLOPE, a technique designed to be user-friendly, can locate and estimate the parameters of hidden taxa. This may open the door for use of taxon search procedures in other substantive areas.

The MAXSLOPE Procedure

We will demonstrate the use of MAXSLOPE on an idealized data set. A more detailed presentation is included in Appendices A and B. The example data set simulates 1000 scores on each of two hypothetical indicator variables, which are shown as a scatterplot in Figure 2a. We deliberately created a large mean separation (about 5 standard deviations) for this example to produce clear-cut, easily visualized results. Scores on the two indicator variables were sets of random numbers converted to integers. In order to mirror the situation with schizotypy, the base rate for the smaller taxon was set at 10%.

The first step in using MAXSLOPE is to perform a procedure called LOWESS regression (Cleveland, 1979). The result of LOWESS regression is a curve that defines the best locally quadratic equation that fits the data. This curve is plotted over the data in Figure 2a. Inspection reveals a clear jump in the shift in the slope near the transition from one taxon to the other. In order to pinpoint the MAXSLOPE, the momentary slope (dy/dx) of the curve is plotted for each point along the *X*-axis as shown in Figure 2b. Because this is a curve of MOMENTARY slope, random variations in the curve can lead to high values for the slope and may make the location of the MAXSLOPE ambiguous. This problem is solved by smoothing the data using a procedure such as SUPERSMOOTH[1], a robust curve smoothing technique (Friedman, 1984). The line overplotted on Figure 2b represents the smoothed slopes of the curve. It is clear from Figure 2b that there is a maximum slope, and therefore a HITMAX, at about *X*=30. Once a HITMAX is located, the base rates of each taxonomic category can be computed using a procedure described in Appendix B.

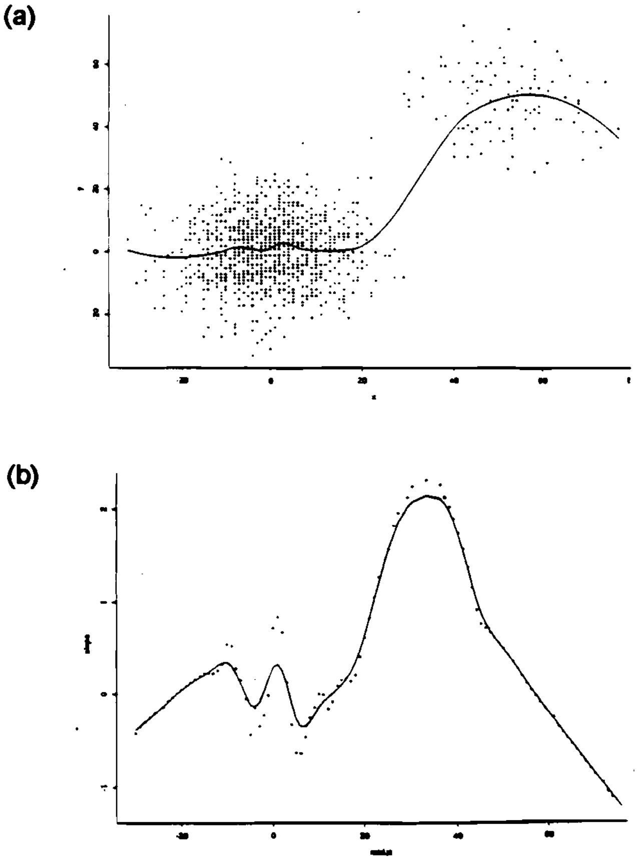

Figure 2 Demonstration of the MAXSLOPE Procedures: (a) Scatterplots on two moderately valid indicators of an underlying taxonomy with local regression line overprinted; (b) Momentary slope of the local regression line from Figure 2a (points represent actual values and smooth line represents the smoothed version of the data).

[1] Note that this is a departure from the procedure outlined by Grove and Meehl (1993). See Appendix A for an explanation of this modification.

Table 1: Summary of the Results for the 18 MAXSLOPE Analyses Using the Cognitive Slippage, Magical Ideation, and Perceptual Aberration Scales.

X-Variable	Y-Variable	Population	HITMAX Found?	Base Rate Estimate	% Above HITMAX
Cog. Slippage	Mag. Ideation	Both Sexes	NO	N/A	N/A
		Males	NO	N/A	N/A
		Females	NO	N/A	N/A
	Percept. Ab.	Both Sexes	YES, CS=18	>1.00	11.2%
		Males	YES, CS=19	.41	11.3%
		Females	YES, CS=20	.20	8.6%
Mag. Ideation	Cog. Slippage	Both Sexes	YES, MI-18	.09	8.6%
		Males	NO	N/A	N/A
		Females	YES, MI=18	.17	9.4%
	Percept. Ab.	Both Sexes	NO	N/A	N/A
		Males	NO	N/A	N/A
		Females	NO	N/A	N/A
Percept. Ab.	Cog. Slippage	Both Sexes	NO	N/A	N/A
		Males	NO	N/A	N/A
		Females	NO	N/A	N/A
	Mag. Ideation	Both Sexes	NO	N/A	N/A
		Males	NO	N/A	N/A
		Females	NO	N/A	N/A

Method

Sample

The subjects were 2877 undergraduates who completed the Perceptual Aberration, Magical Ideation, and Cognitive Slippage Scales as part of a course requirement. Subjects (*N*=286) were dropped if they skipped more than two questions or scored more than two on a five-item Infrequency Scale. There were 1341 males and 1250 females in the final sample.

Implementation

The data were analyzed using a statistical package called *S Plus* by Statsci[2] (Version 4.0 installed on a UNIX Sun Workstation). Quadratic equations used in solving for the base rates were solved using Mathematica on the Sun. The details of the procedures are included in Appendices A and B.

[2] S Plus can be purchased from the publisher, Statistical Sciences, Inc. (STATSCI), 1700 Westlake Ave. N. Suite 500, Seattle WA 98109.

Results

The results were frankly disappointing. Table 1 summarizes the results of the 18 possible MAXSLOPE analyses [each of the three schizotypy scales regressed on each of the other two scales (3X2=6) for males, females, and the combined sample (3)]. When a MAXSLOPE was found, the base rates were estimated using the residuals method suggested by Grove and Meehl (1993).

Discussion

The results of the combined data set are similar to some of the results of previous research. As in Lowrie and Raulin (1990), a HITMAX was found at 18 on Magical Ideation, but regressions on Perceptual Aberration and Cognitive Slippage failed to find a HITMAX. There are several possible explanations as to why a HITMAX was found for Magical Ideation but not Perceptual Aberration. Magical Ideation might be a better indicator of the schizotypic taxon than the others used in this study, and so is the only one powerful enough to show a HITMAX. A second and similar explanation is that the Perceptual Aberration Scale may not be a valid indicator of schizotypy, and so no HITMAX exists. However, this explanation does not explain the positive results of Lenzenweger and Korfine (1992) using this scale. Perhaps the key difference between the current study and the Lenzenweger and Korfine study is the fact that this study used the complete scale while Lenzenweger and Korfine (1992) did itemwise analysis. Perhaps the weaker items on the Perceptual Aberration Scale destroy the predictive validity of the better items, and so what holds for items may not hold for the whole scale. Comparing the distributions of the raw Perceptual Aberration Scale data compared to the distribution of the ideal data suggests a third possibility. The Perceptual Aberration distribution is roughly a line with a positive slope rather than S-shaped like the ideal data. It is possible that there is no MAXSLOPE because the two indicators are too highly correlated within each taxon for the change in slope to be detected. In fact, it is possible that the slope within the taxon is greater than that at the intersection of the taxa. If this were true, one might expect there to be a MINIMUM slope at the HITMAX. Indeed, if one looks at the graph of Perceptual Aberration versus Magical Ideation, one sees a deflection of slope at about $X=14$. This is, however, entirely speculative and any of the possibilities mentioned are plausible.

It is more difficult to understand why this study failed to find a HITMAX for Cognitive Slippage when Lowrie and Raulin (1990) apparently did. One explanation is that the HITMAXes found in their data were artifacts. A second possibility is that MAXCOV procedure may be a more sensitive taxon search procedure. This would suggest that the Cognitive Slippage Scale has a smaller effect size than does Magical Ideation. Finally, it is possible that although not less powerful, MAXSLOPE is less robust to the assumption of zero correlation within taxa. More data are needed to determine whether or not the Cognitive Slippage Scale is sensitive to an underlying schizotypal taxon.

Even more puzzling than the failure to find a HITMAX on Cognitive Slippage is the finding of a sex difference on Magical Ideation. While Lowrie and Raulin (1990) did not find a sex difference, in this study a HITMAX was located for females but not males. The curve for males is nearly bimodal; there are two peaks in the slope. The near-bimodality of the best fit curve has no theoretical explanation. The failure of previous research to find this difference suggests that it might be nothing more than sampling error. A second possibility is that there are somehow three taxa underlying the distribution in males; schizotypes, non-schizotype low scorers, and non-schizotype high scorers. Perhaps some difference in how males and females interpret the questions result in a triple taxonomy for males. The theory upon which taxon search procedures are based does not apply to situations where there are more than two underlying distributions. A third and the most likely scenario is that the sex difference occurred because the sample size for males was too small to accurately pick a small effect size with a small base rate out of background "noise" variation. The reduction in sample size (from about 2500 to about 1250) that occurs when separating by sex will affect the stability of estimates in many steps of the process. Although Monte Carlo studies of the MAXSLOPE procedure (Grove & Meehl, 1993) had no difficulty in running samples of size 1000, they did not test base rates below 25%, whereas here the expected base rate is 10%. Especially if the effect size is small, the low base rate is a significantly more difficult situation. This explanation is especially likely in light of the problems encountered in estimating base rates using the residuals method discussed below.

There is encouraging agreement between the estimates of base rate of the "schizotypic" taxon (9.2%) in the combined sample, Lenzenweger and Korfine's (1992) results (9.8%), and Meehl's (1990) hypothesized size of the schizotypal taxon (10%). The similarity between the results of this study and its predecessors can be taken as convergent validation. Unfortunately, this validation is weakened by the failure to cross-validate these estimates when separated by sex. The estimate of base rate for females at 17% is in the right ballpark but not as convergent as hoped. The failure to find a MAXSLOPE for males is a more serious problem. As suggested above, this result is probably a Type II error caused by small sample size and small effect size. The fact that the base rate estimates for the total sample using the Magical Ideation Scale are so very close to previous research work adds credence to this explanation. However, it appears that further work is needed.

Because MAXSLOPE is a new procedure, there are some important questions that need to be answered. Perhaps a better understanding of these issues would help in understanding the results of this study. One question regarding the use of MAXSLOPE is this author's adaptation to integer data (see Appendix A). The modification is apparently minor and is consistent with the rationale behind the original procedure. Furthermore, the output of both real and ideal data closely resemble output from the original procedure, suggesting that the modification is valid. However, the modification has not been put to empirical test for convergence with the original procedure. Although large sample Monte Carlo data on the success of MAXSLOPE in general are not yet available, the data to date (Grove & Meehl, 1993; as well as this study's results) suggest that, in general, MAXSLOPE is a valid way to identify HITMAX and measure taxonicity.

There are more serious questions about the residuals method of solving for base rates (see Appendix B). The graph of the variance of the residuals in the real data did not result in a graph with the requisite two plateaus. There is a maximum peak, but there are many smaller peaks and valleys. It is not known how robust this method is to variations in the residual graph. Also, Grove and Meehl (1993) state that this method is only exactly correct when the base rates of the two taxa are both 50%. They do not, however, specify how to alter the method when this is not true. The extent to which the 10% base rate results in estimation error is unknown.

In addition to these theoretical problems in using the residual method, some disturbing practical difficulties were encountered. First, there were questions about the width of the sliding cuts. Grove and Meehl (1993) originally set their width at 250 scores, creating about 100 cuts. The authors warn that using too large of a cut to find the variance hides the crucial local character of the variance function and is invalid. This may have happened, for these original base rate estimates led to invalid solutions (base rate estimates of greater than one or less than zero). Also, Grove and Meehl (1993) smoothed the residuals (using Tukey's repeated medians) before solving for base rates. When this was attempted on our data the solution set was again invalid. The solution of 9.2% was found by (1) computing variance in groups of 100 instead of 250, and (2) not correcting for peaks and valleys in the distribution. It is not known how smoothing or not smoothing affects estimation, or if smoothing is even necessary when using integer data as this study does. It appears as if small variations in procedure can lead to large changes in the estimated parameters, and it is far too easy to generate nonsensical or incorrect estimates of base rate. It is important to remember that this problem pertains only to the residuals method of solving for base rates. Other methods, such as one explained in Appendix B that requires three indicators, remain valid, and the validity of the method of estimation is independent of the validity of the taxon search procedure itself.

In summary, the attempted to cross validate previous results regarding the presence of a schizotypal taxon using a new Taxon Search Procedure was only partially successful. Better results were achieved when both sexes were analyzed together, perhaps because of the increase in sample size. As in Lowrie and Raulin (1990), a HITMAX at 18 was found on the Magical Ideation Scale, while no HITMAX was found for Perceptual Aberration. Base rate estimates from the combined sample were 9.2%, comparing favorably to the 9.8% estimate of Lenzenweger and Korfine (1992). However, there were several failures of cross validation as well, particularly when the data were separated by sex. Unlike previous work, no HITMAX was found for males on Magical Ideation. Females showed a HITMAX at 18, on Magical Ideation but the estimate of a base rate of 17% for females is somewhat discrepant from previous estimates. This study also failed to replicate Lowrie and Raulin's (1990) finding of a HITMAX on Cognitive Slippage.

There was partial cross validation of previous results but the results were not as similar as hoped. Summarizing the results of known taxon search studies in schizotypy, Magical Ideation showed underlying taxa in both studies that used it; Cognitive Slippage was found to have underlying taxonicity in one of two studies, and Perceptual Aberration items have shown underlying taxa in one study, but the scale as a whole has not shown taxonicity in two studies. This not-particularly-impressive scorecard is probably due to the fact that this area of research is plagued by small effect sizes and substantial error, which leads to difficulties in finding HITMAXes.

At the same time, it must be remembered that taxon search procedures, being parameter estimation techniques, are far more rigorous mathematical tests of a hypothesis than null hypothesis tests. The record of positive and negative findings cannot be judged by the same standard as in hypothesis tests because there are many more factors that can lead to failure to confirm previous findings. The ability of ANY two studies to replicate similar results is highly unlikely in any parameter estimation study unless there is some validity to the constructs being studied; it is highly unlikely that even the modest cross-validation achieved here is spurious. The failure to replicate some results suggests the possibility that MAXSLOPE may not be as powerful or as robust as MAXCOV, but this needs to be borne out by Monte Carlo simulation. Refinement of the scales themselves, perhaps using taxon search techniques as a guide, promises to increase the predictive validity of our measures. The results of taxon search as well as longitudinal work suggests that this would be a fruitful enterprise, and work in our laboratory along these lines is in the preliminary stages.

On a more positive note, the fact that there was some cross-validation of previous results indicates that MAXSLOPE procedure has promise, even when adapted to use with integers. There are serious questions about the validity of using the residual method of solving for base rates, and until more is known about the procedure, it should be used and interpreted with extreme caution.

This study represents a limited cross-validation of the existence of a schizotypal taxon. It appears that the Magical Ideation Scale is the best indicator of this taxon. MAXSLOPE is a promising new method of exploring data in order to search for the existence of hidden subgroups, although serious questions exist about the residuals method of parameter estimation. More information is needed to add credence to these results, but the current study suggests that pursuing further validation is likely to be fruitful.

References

Chapman, L. J., Chapman, J. P., Kwapil, T. R., Eckblad, M., & Zinser, M. C. (1994). Putatively psychosis-prone subjects 10 years later. *Journal of Abnormal Psychology, 103*, 171-183.

Chapman, L. J.; Chapman, J. P. and Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology, 85*, 374-382.

Chapman, L. J.; Chapman, J. P. and Raulin, M. L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology, 87*, 399 -407.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829-836.

Eckblad, M. and Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology, 51*, 215-225.

Friedman, J. H. (1984). *The variable span smoother.* Technical Report #5, Lab for Computational Statistics, Stanford University.

Furash, L. D; Mahler, C. R.; Raulin, M. L.; O'Gorman, J. C. and Lowrie, G. S. (1988). The relationship between schizotypic signs and symptomatology in a chronic schizophrenic population. Paper presented at the Eastern Psychological Association Convention, Buffalo, April 1988.

Grove, W. M., & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports, 73*, 707-737.

Lenzenweger, M. F. and Korfine, L. (1992). Confirming the latent structure and base rate of schizotypy: A taxometric analysis. *Journal of Abnormal Psychology, 101*, 567-571.

Lowrie, G.S., & Raulin, M. R. (1990). Uncovering a schizotypal taxonomy using mathematical search techniques. Paper presented at the Eastern Psychological Association Convention, Philadelphia, March 1990.

Meehl, P. E. (1962). Schizotaxia, schizotypy, and schizophrenia. *American Psychologist, 17*, 827-838.

Meehl, P. E. (1964). *Manual for use with the checklist for schizotypic signs.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders, 4*, 1-99.

Meehl, P. E. and Yonce, L. J. (1989). *Taxometric analysis by the method of coherent cut kinetics.* Unpublished technical report available from the authors by request.

Miers, T. C. and Raulin, M. L. (1985). The development of a scale to measure cognitive slippage. Paper presented at the Eastern Psychological Association Convention, Boston, March 1985.

Propper, S.; Raulin, M. L.; Lowrie, G. S; Trigoboff, D. H.; Henderson, C. A. and Watson, H. S. (1987). Heterogeneity of schizotypy: Searching for symptom patterns. Paper presented at American Psychological Association Convention, New York, August 1987.

Appendix A
How MAXSLOPE works

MAXSLOPE is designed to analyze a situation such as shown in Figure 1 of this paper [an observable distribution (the outermost curve) is made up of two underlying distributions]. The two underlying distributions represent two populations (taxa). If the taxa cannot be directly measured, as is often the case, taxon search procedure such as MAXSLOPE may allow the parameters of the taxa to be estimated from the mathematical interrelationships of a moderately valid indicators of that taxon.

If you take an interval of a fixed width and slide it on Figure 1 from left to right, the interval will initially have only subjects from the lower taxon. As the interval moves to the right, the subjects in the interval will come from both taxa, and when the interval reaches the point where the two curves cross the proportion of subjects from each taxa will be roughly equal. This point is referred to as the HITMAX location. HITMAX is the decision point on the scale that will maximize the number of correct classifications of subjects. As the interval slides past HITMAX, the majority of subjects defined by the interval will come from the upper taxon.

The behavior of indicator variables around the HITMAX has been exploited by Meehl and his associates in the development of taxonomic search procedures. With a single exception, Meehl's taxon search procedures require at least two and usually more moderately valid indicators of the underlying taxon[3] The exact number varies by procedure. The term moderately valid indicators is used to reinforce the notion that if a perfect indicator of taxon status were available, there would be no need to estimate the parameters of the taxa; they could be measured directly. A typical assumption of Meehl's taxon search procedures is that the moderately valid indicators are uncorrelated within each taxon. Thus, in the scatterplot shown in Figure 2a, there appear to be two circular clusters of scores. Notice that there is a positive correlation between the indicators in the mixed population; this correlation is due, however, to a mean difference between the taxa and not correlations within the taxa. Grove and Meehl (1993) demonstrate that taxon search procedures are in fact somewhat robust to correlations within taxa.

The first taxon search procedure Meehl developed was the Maximum Covariance procedure (MAXCOV). In this procedure, you have three moderately valid indicators that are assumed to be pairwise uncorrelated with each other within each taxonomic category[4]. One of the three variables is designated as the Input Variable and a sliding cut is used on this variable as described above. The covariance of the other two variables is computed within each interval. As you move the sliding cut from right to left, the covariance will increase gradually, peak at the HITMAX, and then fall after the HITMAX. The rationale and details of this procedure are published elsewhere (Meehl, 1973).

There are some practical difficulties with using MAXCOV that are partially handled with the MAXSLOPE procedure. MAXSLOPE has a similar conceptual base as MAXCOV; it relies on the fact that the slope of the regression equation is related to covariance. Looking at the difference score form of the equation for each makes this clear.

Covariance: $s_{xy}=S_{xy}/(N-1)$

Slope: $b=S_{xy}/S_x^2$

The numerators of both equations are the same, while the denominators are not. But consider what happens when one is taking sliding cuts of a distribution. The denominator of the covariance (N-1) is constant; alone, it would plot to a line with a slope of 0. In the regression equation the denominator is S_x^2. When taking sliding cuts of the data, the denominator should slowly increase at a rate proportional to the numerator. Alone, it would plot to a line or a gentle curve. If, on the other hand, there is a local spike in either covariance or slope, it must be due to an increase in the numerator, S_{xy}. Thus, one can expect that some properties of the covariance curve will also be found in the slope. In particular, the slope of the regression line will, like the covariance, be at a maximum in the neighborhood around the HITMAX. Looking at the regression curve plotted over the scatterplot in Figure 2a one can also see this. A steep change in slope occurs in the area between the two taxa. The graph of the slopes themselves (Figure 2b) confirms this. The property of maximum slope means that the HITMAX can be found without needing to find covariance. It still requires a sliding cut of progressive neighborhoods of scores, but one can find the slope instead of the covariance. The primary advantage here is that slopes can be computed from two indicators instead of three. However, another advantage is that the output is easily graphed and familiar to the majority of potential users.

MAXSLOPE uses the locally weighted robust scatterplot smoothing (LOWESS; Cleveland, 1979) procedure to find the best regression solution for the sliding cuts of scores. Because the data are expected to have peaks and valleys rather than be straight, in MAXSLOPE the LOWESS procedure is used to locally regresses the data onto the best-fitting quadratic rather than a line. The output (Y') is not a straight line, but rather reflects local variations in scores. LOWESS is used instead of other potential procedures because it is designed to be robust to outliers (Cleveland, 1979). The maximum slope occurs at HITMAX, and this usually cannot be determined merely by looking at the regression curve. Therefore, we determine and graph the momentary slope of the regression line at each point (Figure 2b). This is computed as follows. The slope is defined as the change in Y over the change in X. We can thus find the difference of each predicted point from the point before it on X and on Y. Their quotient is the slope of the curve at that point. We graph the values of the slope over the range of the data, producing a dy/dx graph. These dy/dx graphs typically show a lot of random variation. Consequently, the MAXSLOPE is not always unambiguous. However, unlike other points, there should be a general

[3] The exception, called the Normal Single Indicator Method, has several disadvantages that make it a poor choice of Taxon Search procedure. Instead of requiring multiple indicators it assumes normality in the taxa, and unpublished research by this author suggests it may not be very robust. Further, its output--a matrix of chi-squares--is extremely difficult to interpret.

[4] This assumption was needed to derive the mathematics of the MAXCOV procedure, but Monte Carlo studies suggest that the procedure is quite robust to violations of this assumption.

increase of slope in several successive neighborhoods leading up to the MAXSLOPE, as opposed to a single spike caused by error. To find this point, the curve is smoothed using any one of many procedures, such as Tukey's repeated medians or SUPERSMOOTH (Freidman, 1984).

Grove and Meehl (1993), in their paper introducing the MAXSLOPE procedure, suggest doing a second LOWESS regression in order to smooth out the data and help ensure proper location of the MAXSLOPE. They use the following procedure. They start by generously bracketing off the general area where the MAXSLOPE is located (this should be obvious) and then repeat the LOWESS a second time on this area. This procedure should pinpoint the MAXSLOPE. However, this approach does not work when working with integer data, as is the case with the schizotypy scales. Therefore, another approach is needed. The problem lies in the finding of momentary slope by dy/dx. With real data, NO two data points are EXACTLY the same. If there are 1000 data points, there will be 1000 slightly different values of X and consequently of Y'. Thus, dy/dx is some real number at all values, and LOWESS can be performed. With integer data, MANY data points are exactly the same. For instance, the Perceptual Aberration Scale has a range of 0-34. As a result, although there may be 1000 data points, the slope is ONLY DEFINED AT THE THIRTY OR SO POINTS IN THE DISTRIBUTION WHERE THE X VALUES CHANGE. Thus, if there are forty scores of 1 on scale X, dy/dx will be defined for the first of these; X changes from 0 to 1, and Y' changes from predicted value at 0 to the predicted value at X=1. But for the second point where X=1, there is no change in X (one minus one is zero) and no change in the predicted value of Y because X is the same. The slope dy/dx will therefore be undefined (0/0) at these other points. These undefined values of dy/dx make it impossible to perform the second LOWESS on integer data as was done in Grove and Meehl (1993).

Fortunately, the inability to perform a second LOWESS on integer data does not cause any significant problems for the procedure. This is because with inspection it becomes apparent that the second regression is meant to iron out microvariations on the dy/dx plot. However, the plot of dy/dx has thirty or so data points instead of 1000--there ARE no microvariations to smooth. Thus, it appears to be sufficient to substitute a less powerful smoothing technique such as Tukey's repeated medians or SUPERSMOOTH (Freidman, 1984) when using integer data.

Regardless of how it is done, after the plot of the slopes has been smoothed, there should be a clear maximum local slope. This is the HITMAX. The highly visual nature of MAXSLOPE is its the primary advantage. However, in addition to knowing if the data suggest taxonicity, it is desirable to know something about the taxa, such as there base rates. Methods for estimating base rates are covered in Appendix B.

Appendix B
Solving for Base Rates with MAXSLOPE

There are at least two ways to solve for the base rates of the two taxa underlying the observed distribution using MAXSLOPE. The simpler and less problematic requires at least three indicators. The more difficult has the advantage of only requiring two measures, but there are problems with this procedure that suggest the need for further study and validation.

The simpler way to solve for base rates is to create a system of equations and solve algebraically. Meehl (1973) showed that if one assumes that test scores are uncorrelated in each taxon, then it is true that

$\max s_{xy}=\frac{1}{4}D_xD_y$.

If you have three indicators, then you can find the covariances between all three and solve a system of three equations and three unknowns.

$\max s_{xy}=\frac{1}{4} D_xD_z$

$\max s_{yz}=\frac{1}{4}D_xD_z$

$\max s_{xz}=\frac{1}{4}D_xD_z$.

The method of solving for base rates in MAXCOV-HITMAX is derived from these same equations, but this method is unique to MAXSLOPE because only MAXSLOPE yields estimates of D_x, D_y and D_z.

Unfortunately, more than two valid indicators are seldom available. Recall that the principal advantage of MAXSLOPE is precisely the fact that only two indicators are needed. Grove and Meehl (1993) outline another way to estimate base rates using just two indicators. We will refer to this method as the residuals method for reasons that will become obvious shortly. The center of this method is the following equation:

$s^2_X=Ps^2_{LX} + Qs^2_{HX} + PQD^2_X$.

where P and Q are the base rates of taxon 1 and 2, respectively. In words, the variance of the indicator is equal to sum of the variance of the indicator within each population multiplied by the size of that population plus the product of the sizes of the taxa times a factor of rate of change. The variance of the indicator is straightforwardly arrived at, but in order to solve for base rates, P and Q, the other parameters of the model must be estimated. Grove and Meehl (1993) present the following method for arriving at estimates of these parameters. First, a graph is prepared of the squared residuals between the original data and that predicted by the LOWESS curve (see Appendix A). Then, sliding cuts are taken of the squared residual data and the variance of the residuals is found. The expected result is a curve that has two plateaus and a peak in the middle. The plateaus represent s^2_{LX} and s^2_{HX}, respectively. When P=Q, then D^2_X = the height of the peak minus the height of each plateau. With the estimates and the fact that P=1-Q, a quadratic equation can be solved to arrive at estimates of the base rate P.

In using this method to arrive at estimates of base rate, potential problems were uncovered that warrant further research. First, Grove and Meehl (1993) suggest that if the base rates are not equal, other ways of estimating D^2_X may be required, but they do not explain how that might be done. Second, it was found that a plateau is not always found, but rather an irregular graph might be found on one or the other side. Attempting to smooth this plateau led to an impossible solution, but leaving it be led to a very plausible answer (see Results). Third, they suggest that the peak of the squared residuals should be at the MAXSLOPE; it is not clear if this is necessary, and if it is even possible when the base rate is very small; data points are lost in this process, and cuts whose midpoint were at MAXSLOPE might not exist. Finally, there is some question about the nature of the sliding cut. Grove and Meehl (1993) suggest that you can decide the size of the cut either by selecting fixed-N window or selecting N such that a fixed number of windows are created. Beginning with the latter approach, a window size of N=250 (10% of the population) was selected. This led to yet another impossible solution set. Changing the width of the window to N=100 (4%) resulted in a plausible solution. These results suggest that this residual method of estimating base rates might not be particularly robust to certain parameters of estimation. This latter method of estimating base rates warrants further study. These problems suggest that this method of solving for P is less valid than MAXSLOPE itself. Monte Carlo studies of this procedure are needed to assess its accuracy.