

A Bayesian Approach to Reducing Heterogeneity in Laboratory Measures: An Illustration from Schizophrenia Research

Shane T. Jensen
Mark F. Lenzenweger
Donald B. Rubin

ABSTRACT Heterogeneity in the performance of persons affected with schizophrenia or schizotypy psychopathology on laboratory tasks has long been recognized for the challenges it poses for experimental psychopathology, genetic, and other investigations. Traditional techniques such as factor analysis, discriminant function analysis, and cluster analysis have all been deemed inadequate for resolving heterogeneity due to one or another statistical shortcoming or limitation. A group of experimental subjects was initially identified as schizotypic using the well-known Perceptual Aberration Scale. We present a Bayesian approach, computationally implemented using a Gibbs sampling strategy, that enables one to effectively parse this experimental group in a manner that reduces heterogeneity and allows for the separation of what are termed true schizotypes and false-positives. This approach is complemented by maximum likelihood estimation, based on the expectation-maximization (EM) algorithm. The validity of our parsing strategy is supported by reference to other laboratory indexes of relevance to schizophrenia and schizotypy that were not included in the initial analyses.

1 Introduction

1.1 Psychological Background

Schizophrenia, a profoundly disabling psychotic illness affecting approximately 1% of the world's population, has long been known to be characterized by considerable heterogeneity in its clinical presentation (Bleuler, 1911/1950; Kraepelin, 1919/1971). Laboratory investigations of schizophrenia have encountered the effects of heterogeneity, forcing many researchers to focus on the development of heterogeneity-reducing strategies (e.g., Chap-

man and Chapman, 1977, 1989). For example, although there are some well established laboratory findings about the illness e.g., eye tracking dysfunction (Levy et al, 1993), deficits in sustained attention (Cornblatt and Keilp, 1994), most schizophrenia cases do not display more than one of these deficits. The general effect of heterogeneity in any investigation of schizophrenia has been increased noise and, thereby, obscured signal.

The study of schizophrenia and its manifestations is further complicated by the fact that the illness picture is blurred by additional confounding factors such as medication, institutionalization, deterioration, and stigma of diagnosis effects. One way to attempt to deal with this issue has been to study *schizotypes*, those persons “at risk” for the illness. At-risk individuals are currently unaffected with schizophrenia but can display a milder form of some schizophrenic symptoms, such as delusions, hallucinations or social withdrawal. Schizotype individuals do not present the complicating features of schizophrenics such as medication, institutionalization, deterioration, and stigma of diagnosis effects. However, heterogeneity still exists in the laboratory performance of groups known to be at-risk for the illness (Lenzenweger, 1998).

1.2 Classification and Laboratory Measures

Perceptual Aberration Scale (PAS).

The PAS is a 35-item true-false self-report measure of disturbances and distortions in perceptions of body image as well as of other objects (Chapman et al, 1978). A corpus of data support the validity of the PAS as a valid, though imperfect, psychometric indicator of some aspects of schizotypy (Chapman et al, 1995; Lenzenweger, 1998).

Failure to Maintain Set (FMS).

A computerized Wisconsin Card Sorting Test (WCST) was administered according to the standard guidelines specified in the WCST manual (Heaton, 1981) and scored using a computerized version of the test (Harris, 1988). The WCST is the well-known neuropsychological task that measures abstraction ability and cognitive flexibility. For the purposes of this analysis, we used the WCST performance index known as “failure to maintain set” (FMS), which assesses loss of the correct sorting principle needed to perform the WCST properly (Harris, 1998). Prior work has shown that the FMS index discriminates schizotypes from non-schizotypes (Lenzenweger and Korfine, 1994; Park et al, 1995). Higher FMS scores are indicative of poorer performance on the WCST.

Eye Movement Measurement (ETD).

Smooth pursuit eye movement was recorded using a high-speed infrared eye tracking system and the details for this assessment are given in O’Driscoll (1998). Eye movement pursuit performance was evaluated independently

by two expert raters blind to subject identity and group membership. The mean of the two raters served as the basic performance index for the eye tracking performance. As with FMS, ETD scores are coded in the analyses such that higher scores are indicative of poorer performance.

Classification based on the PSA measure.

Separate group means and standard deviations for males and females on the PAS test were computed and served as the basis for initial classification. Following (Chapman and Chapman, 1985), subjects were required to have scored at least 2.0 standard deviations above the group mean on the PAS, whereas normal controls were required to have scored no higher than 0.5 standard deviations above the group mean. Study subjects for each of the two groups were selected at random from the two sub-samples of subjects meeting the specified criteria. The dataset consists of 21 classified normal and 25 initially classified schizotype individuals. Differences between the schizotype and normal study groups were minimal with respect to gender proportion or agreement to participate in the study. Subsequent laboratory testing for FMS and ETD was carried out blind to the group membership.

1.3 Objective

Traditional statistical approaches have been unable to adequately address the heterogeneity of task performances in schizophrenia research. For example, factor analytic procedures help to reduce large numbers of variables to a smaller set of factors, discriminant analysis seeks linear combinations of variables that separate diagnostic groups, profile analysis seeks to determine if groups differ in their configuration of performance on variables or measures of interest. None of these methods, however can adequately separate individuals who might represent true schizotypes from those who represent noise within a classified schizotype group, nor do their various computations take into account adequately the heterogeneity in performance of normal subjects. Finally, we note that although cluster analysis seeks to parse subjects into meaningful subclasses and can aid in an investigation of latent structures, the marked limitations of cluster analysis for even simple parsing tasks (e.g., male vs. female) are known (Golden and Meehl, 1980) and limit enthusiasm for the technique in studies of heterogeneity. Moreover, cluster analysis lacks a formal statistical procedure for the determination of the proper (or correct) number of classes underlying a multivariate space.

Therefore, we undertook the present study in an effort to utilize a multivariate database gathered from individuals who were identified as putative schizotypes or non-schizotypes and a statistical approach based on a Bayesian mixture model designed to sift through the laboratory task performance patterns of these study subjects in order to reduce heterogeneity among the schizotypes. In short, this analysis sought to segregate those

classified schizotypic subjects who are most likely to be true schizotypes from those who are most likely to be false positive cases.

2 Mixture Modelling

Since FMS and ETD are both discrete measures, each individual can be represented by a single count in a contingency table with dimensions defined by their possible outcomes on the FMS ($i = 1 \dots 6$) and ETD ($j = 1 \dots 8$) tasks. We have separate two-dimensional tables of cell counts for classified normals N_{ij} and classified schizotypes S_{ij} . Our mixture model makes the following assumptions:

1. The classification is correct for all individuals classified as normal.
2. The classification is potentially incorrect for individuals initially classified as schizotypes ie. some of those classified as schizotypes are false positives.
3. Normal individuals misclassified as schizotypes have the same model on their observed performance measures as normal individuals who were correctly specified
4. FMS and ETD measures are assumed to be independent in normal individuals, whereas no independence is assumed for the true schizotype group (the actual correlation between the FMS and ETD scores in the normals was only .13 and reasonably consistent with this assumption).

Thus, we are assuming that the classified schizotype group is actually a mixture of true schizotypes and true normals, whereas all individuals in the classified normal group are true normals. A graphical representation of our mixture model is shown in Figure 1.

Our fourth assumption refers to the underlying probability model on the counts in the true normal N_{ij}^* and true schizotype S_{ij}^* table:

$$\begin{aligned} N_{ij}^* &\sim \text{Multinomial}(\pi_{ij}^N), \text{ with an Independence model on } \pi_{ij}^N, \text{ and} \\ S_{ij}^* &\sim \text{Multinomial}(\pi_{ij}^S), \text{ with a Saturated model on } \pi_{ij}^S. \end{aligned}$$

Our parameters of interest are λ , the proportion of true schizotypes among classified schizotypes, and $\boldsymbol{\pi}^N, \boldsymbol{\pi}^S$, the cell probabilities for the true normal and true schizotype tables.

The Bayesian approach to this problem considers the unknown parameters as random variables with a prior distribution, which can be combined with the observed likelihood to form a posterior distribution. In this context, it is beneficial to obtain the posterior distribution of the parameters

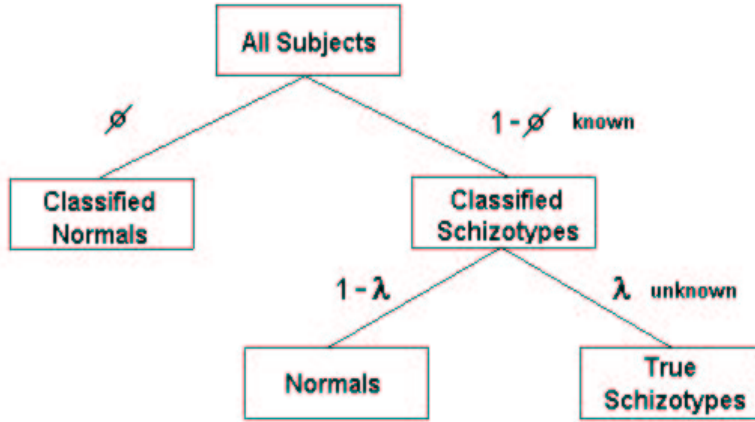


FIGURE 1. Mixture Model for Classified Schizotypes

instead of relying solely on a point estimate, such as the maximum likelihood estimate generated by an EM algorithm:

$$\begin{aligned} \text{posterior} &\propto \text{likelihood} \cdot \text{prior} \\ p(\lambda, \pi^N, \pi^S | \mathbf{N}, \mathbf{S}) &\propto p(\mathbf{N}, \mathbf{S} | \lambda, \pi^N, \pi^S) \cdot p(\lambda, \pi^N, \pi^S) \end{aligned}$$

However, the posterior distribution $p(\lambda, \pi^N, \pi^S | \mathbf{N}, \mathbf{S})$ does not have a simple analytical form, since our underlying multinomial model is on the true counts \mathbf{N}^* and \mathbf{S}^* , not our classified counts \mathbf{N} and \mathbf{S} .

Instead, our strategy is to augment our observed data with \mathbf{I} , a vector of missing indicator variables for true schizotypal status for each individual in the classified schizotypal group.

$$I_k = \begin{cases} 1 & \text{if } k\text{-th subject is a true schizotypal} \\ 0 & \text{if } k\text{-th subject is a true normal} \end{cases}$$

The Gibbs sampler (Geman and Geman, 1984) can be used to obtain draws that converge to the joint posterior distribution $p(\lambda, \pi^N, \pi^S, \mathbf{I} | \mathbf{N}, \mathbf{S})$. Given initial values of all the parameters, the Gibbs sampler makes successive draws from the conditional distributions of one set of parameters, conditioned on all other parameters, in our case:

1. $p(\mathbf{I} | \pi^S, \pi^N, \lambda, \mathbf{N}, \mathbf{S})$
2. $p(\lambda | \pi^S, \pi^N, \mathbf{I}, \mathbf{N}, \mathbf{S})$
3. $p(\pi^S | \lambda, \mathbf{I}, \mathbf{N}, \mathbf{S})$
4. $p(\pi^N | \lambda, \mathbf{I}, \mathbf{N}, \mathbf{S})$

From a simple application of Bayes rule, for the k -th subject who is located in the (i, j) cell of the schizotype table,

$$I_k \mid \pi_{ij}^S, \pi_{ij}^N, \lambda, \mathbf{N}, \mathbf{S} \sim \text{Bernoulli}(p),$$

where, with the known proportion of initially classified schizotypes ϕ ,

$$p = \frac{\pi_{ij}^S \cdot (1 - \phi)\lambda}{\pi_{ij}^S \cdot (1 - \phi)\lambda + \pi_{ij}^N \cdot (\phi + (1 - \phi)(1 - \lambda))}.$$

Given a drawn indicator variable I_k for each subject k in the classified schizotype group, we can augment our data to get the true counts N_{ij}^* and S_{ij}^* .

With a vague prior distribution on λ , $\lambda \sim \text{Unif}(0, 1)$, we have for the conditional posterior distribution of λ

$$\lambda \mid \mathbf{I}, \mathbf{N}, \mathbf{S} \sim \text{Beta} \left(\sum_{ij} S_{ij}^* + 1, \sum_{ij} S_{ij} - \sum_{ij} S_{ij}^* + 1 \right).$$

For π^S with a fully-saturated structure and a vague ($\theta_{ij} \rightarrow 0$) conjugate prior distribution $\pi_{ij}^S \sim \text{Dirichlet}(\theta_{11}, \dots, \theta_{IJ})$, we have

$$\pi_{ij}^S \mid \mathbf{I}, \mathbf{S} \sim \text{Dirichlet}(S_{11}^* + \theta_{11}, \dots, S_{IJ}^* + \theta_{IJ}).$$

For π^N with an independence structure and the same vague conjugate prior distribution as π^S , we take $\pi_{ij}^N \mid \mathbf{I}, \mathbf{N}$ as the product $\pi_{ij}^N = \alpha_{i+} \cdot \beta_{+j}$ of marginal probabilities

$$\begin{aligned} \alpha_{i+} \mid \mathbf{I}, \mathbf{N} &\sim \text{Dirichlet}(N_{1+}^* + \theta_{1+}, \dots, N_{I+}^* + \theta_{I+}) \text{ and} \\ \beta_{+j} \mid \mathbf{I}, \mathbf{N} &\sim \text{Dirichlet}(N_{+1}^* + \theta_{+1}, \dots, N_{+J}^* + \theta_{+J}). \end{aligned}$$

3 Results

From the resulting draws from the joint posterior distribution, means, standard deviations, and 95% intervals were calculated for each parameter. A subset of these parameters are included in Table 1. All cell probabilities included in Table 1 are for cells with positive counts in the observed data.

The posterior interval for λ suggests that a substantial proportion of individuals in the classified schizotype group are not actually true schizotypes, but instead should be considered false positives.

We can see that for some cells in the classified schizotype table that initially contained positive counts, posterior cell probabilities have dropped down towards zero in the true schizotype table. The probability mass in these cells has been moved to the true normal table, resulting in a change in the corresponding true normal cell probabilities.

Parameter	Mean	Standard Deviation	95% Posterior Intervals
π_{11}^N	0.2238	0.0567	(0.1297 , 0.3443)
π_{12}^N	0.0733	0.0342	(0.0240 , 0.1618)
π_{15}^N	0.1037	0.0374	(0.0471 , 0.1827)
π_{21}^N	0.1512	0.0435	(0.0749 , 0.2415)
π_{28}^N	0.0803	0.0308	(0.0304 , 0.1516)
π_{38}^N	0.0236	0.0134	(0.0055 , 0.0561)
π_{11}^S	0.0000	0.0000	(0.0000 , 0.0000)
π_{12}^S	0.0000	0.0000	(0.0000 , 0.0000)
π_{15}^S	0.0000	0.0000	(0.0000 , 0.0000)
π_{23}^S	0.2282	0.1279	(0.0301 , 0.5217)
π_{38}^S	0.0000	0.0006	(0.0000 , 0.0000)
π_{46}^S	0.1182	0.1030	(0.0034 , 0.3849)
π_{57}^S	0.1061	0.0947	(0.0037 , 0.3480)
π_{63}^S	0.1115	0.0966	(0.0044 , 0.3567)
λ	0.3847	0.0949	(0.2108 , 0.5730)

TABLE 1. Summary of Posterior Draws

Our model formulation and assumption that not all classified schizotypes are likely to be true schizotypes has resulted in a dramatic change in the schizotype table between the empirical (shown in Table 2) and posterior cell probabilities.

Our results are complemented by the following table of maximum likelihood estimates for the cell probabilities π^N and π^S . These estimates were calculated with an Expectation-Maximization (EM) algorithm (Dempster et al, 1977) that utilized the same missing data framework as our Bayesian analysis (details can be found in Lenzenweger et al, 2001).

Confirming the results based on our Bayesian formulation, significant differences can be observed between the classified schizotypic probabilities and the maximum likelihood estimates of the true schizotypic probabilities, with the proportions in several cells of the true schizotype table decreasing towards zero as a result of these individuals being moved to the true normal table.

Empirical Probabilities

	Classified Normals					Classified Schizotypes							
	1	2	5	7	8	1	2	3	4	5	6	7	8
1	0.38	0.05	0.05		0.10	0.08	0.08		0.04	0.04	0.04		0.08
2	0.14	0.05	0.05	0.05	0.05	0.08		0.08		0.08	0.04		0.08
3		0.05	0.05							0.04			0.04
4											0.04		
5												0.04	0.04
6								0.04					

Maximum Likelihood Estimates

	True Normals					True Schizotypes							
	1	2	5	7	8	1	2	3	4	5	6	7	8
1	0.21	0.07	0.09	0.01	0.13				0.10		0.10		
2	0.16	0.05	0.07	0.01	0.09			0.21			0.10		
3	0.04	0.01	0.02	0.00	0.02					0.04			0.02
4											0.10		
5												0.10	0.10
6								0.10					

Note: columns are categories of ETD performance and rows are categories of FMS performance

Table 2: ML estimates from an EM algorithm

Finally, our posterior draws of the missing data vector \mathbf{I} were used to calculate the probability of each classified schizotype being a true schizotype versus a true normal. To begin, we note that the posterior probabilities for almost all individuals were either essentially zero or near to one. This clearcut separation of the probabilities allowed us to unambiguously parse the classified schizotypes into either a true schizotype or a false positive group. This parsing yielded 10 true schizotypes and 15 false positives from the 25 originally classified schizotypes.

Several other laboratory measures that were not included in the model were examined under the new classification. In addition to the FMS and ETD measures, CPT-IP reaction time, total thought disorder (TDI) and delayed response task measures were also collected from each individual in the study. CPT-IP reaction time is a measurement of sustained attention involving reaction times in a continuous performance test. Total thought disorder is a coded measure of thought disordered responses in a verbal description of a visual stimulus. Delayed response task measures spatial working memory of a visual stimulus. More details on these measures can be found in Lenzenweger et al (2001). Table 3 gives the mean and standard deviation of performance on these three measures with our new classification of true schizotypes, true normals, and false positives.

Measure	True Normals		False Positives		True Schizotypes	
	Mean	SD	Mean	SD	Mean	SD
CPT-IP Reaction Time	543.53	53.69	572.48	60.18	595.44	59.19
Delayed Response Task	93.26	5.40	90.85	7.95	86.65	8.70
Total Thought Disorder	2.57	3.17	5.79	8.27	9.50	15.01

TABLE 3. Other task performances with revised classification

All three measures are consistent in showing that the performance of normal individuals and true schizotypes are noticeably different, with the false-positive group performing somewhere in between these two extremes.

These results are not ideal in terms of our model assumptions, since normal and false positive individuals do not display the same performance, but these results do suggest that our refined classification has reduced some of the heterogeneity in the classified schizotype group on these performance measures. This is an especially encouraging result since these performance measures were not considered in our model and therefore could not directly influence our refined classification.

4 Convergence and Sensitivity to Priors

The convergence of the Gibbs Sampler to the desired posterior distribution can be investigated by running multiple chains, each from different well-dispersed starting values and examining whether or not the chains move together to the same distribution of values. These chains can be examined analytically by looking at the “shrink factor” \hat{R} proposed by Gelman and Rubin (1992) for each parameter, which compares the between-chain variance to the within-chain variance. Using well-dispersed starting values for fifteen multiple chains, \hat{R} after 2000 iterations was calculated to be close to 1 for each parameter.

Although a converged Gibbs sampler will produce the desired draws of each parameter from their joint posterior distribution, the draws are not independent. Autocorrelation plots of each parameter in each chain were examined, and all seemed to show very little correlation. A sample of these autocorrelation plots for the selected parameters $\lambda, \pi_{11}^N, \pi_{12}^N, \pi_{23}^S$ are shown in Figure 2.

Several different prior distributions were also implemented for π^N, π^S , and λ , but none led to posterior intervals that were substantially different from those summarized in Table 1.

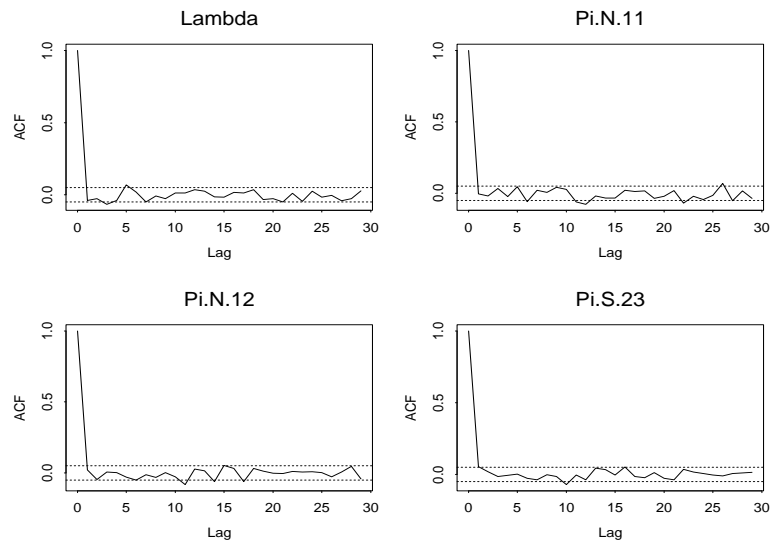


FIGURE 2. Autocorrelation functions of selected parameters

5 Discussion

The objective here was to apply a Bayesian mixture model to empirical data collected on two laboratory measures, FMS and ETD, in an effort to reduce heterogeneity in performance on those measures among the initially classified schizotypal subjects. The results seem to indicate that the initially classified schizotypal group contained a large proportion of individuals that were not true schizotypes.

Our analysis also allowed us to parse each individual in the classified schizotypal group into a true vs. false positive schizotypal dichotomy, where this new classification appears to have validity based on several other performance measures not included in the mixture model.

Thus, our approach enabled us to go into the classified schizotypal group and objectively divide subjects into two groups which possess different performance across a number of laboratory measures, even though their phenotypic presentation (PAS score) did not demarcate them from each other. This, we argue, represents a potential methodological advance for this area of schizotypal and schizophrenia research.

Acknowledgments

We thank Michael J. Coleman, Philip S. Holzman, Deborah L. Levy, Gillian O'Driscoll, and Sohee Park for their collaborative efforts with the second author on the original studies that provided portions of the data for this illustration.

References

- Bleuler, E. (1911/1950). *Dementia praecox or the group of schizophrenias*. (J. Zinkin, Trans.). New York: International Universities Press.
- Chapman, L.J., and Chapman, J.P. (1977). Selection of subjects in studies of schizophrenic cognition, *Journal of Abnormal Psychology* **86**, 10-15.
- Chapman, L.J., and Chapman, J.P. (1985). Psychosis proneness. In M. Alpert (Ed.), *Controversies in schizophrenia: Changes and constancies* (pp. 157-172). New York: Guilford Press.
- Chapman, L.J., and Chapman, J.P. (1989). Strategies for resolving the heterogeneity of schizophrenics and their relatives using cognitive measures, *Journal of Abnormal Psychology* **98**, 357-366.
- Chapman, J.P., Chapman, L.J., and Kwapil, T.R. (1995). Scales for the measurement of schizotypy. In A. Raine, T. Lencz, & S. Mednick, (Eds.), *Schizotypal personality* (pp. 79-106). New York: Cambridge University Press.
- Chapman, L.J., Chapman, J.P., and Raulin, M.L. (1978). Body-image aberration in schizophrenia, *Journal of Abnormal Psychology* **87**, 399-407.
- Cornblatt, B.A., and Keilp, J.G. (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia, *Schizophrenia Bulletin* **20**, 31-46.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**, 1-38.
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science* **7**, 457-511.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.
- Golden, R. and Meehl, P. (1980). Detection of biological sex: an empirical test of cluster methods, *Multivariate Behavioral Research* **15**, 475-496.
- Harris, M.E. (1988). *Wisconsin Card Sorting Test: computer version, research edition*. Odessa, FL: Psychological Assessment Resources, Inc.
- Heaton, R.K. (1981). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Kraepelin, E. (1919/1971). *Dementia praecox and paraphrenia* (R.M. Barclay, Trans: G.M. Robertson, Ed.). Huntington, NY: Krieger. (Original work published 1909-1913; original translation of selected portions published 1919).

- Lenzenweger, M.F. (1998). Schizotypy and schizotypic psychopathology: Mapping an alternative expression of schizophrenia liability. In M.F. Lenzenweger & R.H. Dworkin (Eds.), *Origins and development of schizophrenia: Advances in experimental psychopathology* (pp. 93-121). Washington, DC: American Psychological Association.
- Lenzenweger, M.F., Jensen, S.T., and Rubin, D.B. (2001). Finding the "true" schizotypic: On resolving heterogeneity in performance on biobehavioral laboratory measures. Submitted to *Journal of Abnormal Psychology*.
- Lenzenweger, M.F., and Korfine, L. (1994). Perceptual aberrations, schizotypy and the Wisconsin Card Sorting Test. *Schizophrenia Bulletin* **20**, 345-357.
- Lenzenweger, M.F., Loranger, A.W., Korfine, L., and Neff, C. (1997). Detecting personality disorders in a nonclinical population: Application of a two-stage procedure for case identification. *Archives of General Psychiatry* **54**, 345-351.
- Levy, D.L., Holzman, P.S., Matthysse, S., and Mendell, R. (1993). Eye tracking dysfunction and schizophrenia: A critical perspective. *Schizophrenia Bulletin* **19**, 461-536.
- O'Driscoll, G., Lenzenweger, M.F., and Holzman, P.S. (1998). Antisaccades and smooth pursuit eye tracking and schizotypy. *Archives of General Psychiatry* **55**, 837-843.
- Park, S., Holzman, P.S., and Lenzenweger, M.F. (1995). Individual differences in working memory in relation to schizotypy. *Journal of Abnormal Psychology* **104**, 355-363.