ON TESTING TWO THEORIES REGARDING THE GENETIC MAKEUP OF PATIENT--ETCU

SEP 81 T ROBERTSON, G WARRACK

UNCLASSIFIED TR-76

END
ON TESTING TWO THEORIES REGARDING THE GENETIC MAKEUP
OF PATIENTS SUFFERING FROM UNIPOLAR AFFECTIVE DISORDER

Tim Robertson and Giles Warrack

Department of Statistics
The University of Iowa
Iowa City, Iowa 52242

Technical Report #76
September 1981
ON TESTING TWO THEORIES REGARDING THE GENETIC MAKEUP OF PATIENTS SUFFERING FROM UNIPOLAR AFFECTIVE DISORDER*

Tim Robertson and Giles Warrack

SUMMARY

In an article in the Journal of Psychiatric Research, Cadoret, Woolson and Winokur (1977) consider two theories regarding the genetic makeup of patients suffering from unipolar affective disorder. Using likelihood ratio techniques from order restricted inference we analyze their data and compare the two analyses.

*This research was sponsored by the Office of Naval Research under Contract N00014-80-C-0321.
INTRODUCTION. In an article in the Journal of Psychiatric Research, Cadoret, Woolson and Winokur (1977) consider the genetic factors contributing to the risk of unipolar affective disorders (u.a.d.). They consider two theories accounting for the age of onset of u.a.d.: (1) the "qualitative" theory, which postulates that the genetic makeup of those afflicted early in life (under 40 years old), and those who are afflicted later (over 40) is somehow different, and 2) the "quantitative" theory which maintains that there is a particular type of gene causing u.a.d., and that those who succumb earlier do so because they possess the gene in greater numbers.

In their investigation Cadoret et al. consider samples of 767 women and 398 men suffering from u.a.d. Each sample is divided into 6 age groupings (3 less than 40 and 3 greater than 40) according to the age at which the patient first suffered from the disorder. For each age group they have obtained (a) the proportion of patients having alcoholic fathers, and (b) the proportion of fathers and mothers of patients within the group who suffered from depression.

These data reflect on the genetic makeup of the families in question and the two above-mentioned theories are modeled in terms of the shapes of the functions relating the probability of alcoholism or depression in parents to the age of onset of the patients' illness. If the "qualitative" theory holds then the risk function should be constant until age 40, drop at age 40, and be constant thereafter. If the "quantitative" theory holds
then the risk function should simply be nonincreasing. The shapes of the risk function under the two theories are shown in Figure 1.

![Figure 1. Shapes of the risk function under the two theories explaining the age of onset of u.a.d.](image)

The six age groupings are 0-20 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, and over 60 years. Let $p_i$, $i=1, 2, \ldots, 6$ be the respective risks for these age groups (probability of an alcoholic father is one case and the probability of a parent suffering from depression in the other case).

In terms of statistical hypotheses, the qualitative theory implies that $H_0: p_1 = p_2 = p_3 \geq p_4 = p_5 = p_6$ while the quantitative theory implies that $H_A: p_1 \geq p_2 \geq \cdots \geq p_6$. Note that $H_0$ implies $H_A$ so that an appealing way to decide between the two theories would be to use the data to test $H_0$ as a null hypothesis when the alternative is restricted by $H_A$ (i.e., test $H_0$ vs. $H_A - H_0$ ($H_A$ but not $H_0$)).
The data used by Cadoret et al. (1977) is contained in Table 1. In analyzing this data they use the techniques of order restricted inference contained in Barlow, Batholomew, Bremner and Brunk (1972). However, distribution theory was not available for testing $H_0$ against $H_A - H_0$ and Cadoret et al. tested separately the hypotheses

(a) $H'_0: p_1 = p_2 = p_3$ vs. $H'_A - H'_0$ where $H'_A: p_1 \geq p_2 \geq p_3$

and

(b) $H'_0: p_4 = p_5 = p_6$ vs. $H'_A - H'_0$ where $H'_A: p_4 \geq p_5 \geq p_6$.

In each case the test statistic is of the form $-2 \ln \Lambda$, where $\Lambda$ is the likelihood ratio and the asymptotic distribution is the chi-bar-square contained in Theorem 3.1 of Barlow et al. Thus for each set of data they computed two $P$-values. These $P$-values together with the results of our analysis (described later) are given in Table 2.

There are a number of techniques for combining $P$-values from independent tests. One of the more appealing techniques is based upon the facts that the null hypothesis distribution of a $P$-value is uniformly distributed over $(0,1)$ and that $-2 \ln(U)$ ($U$ is distributed uniformly on the interval $(0,1)$) has a chi-square distribution with two degrees of freedom. Thus in Table 2 we have combined the two $P$-values from the independent tests by computing $P[X^2_4 \geq -2(\ln p_1 + \ln p_2)]$, where $p_1$ and $p_2$ are the $P$-values obtained for the two separate tests.
TABLE 1. Proportions of alcoholic fathers and parents suffering from depression

A. FEMALES

<table>
<thead>
<tr>
<th>Age at onset of illness</th>
<th>Sample Size</th>
<th>Proportion of alcoholic fathers</th>
<th>Proportion of depressive parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 20 yrs.</td>
<td>68</td>
<td>.16</td>
<td>.16</td>
</tr>
<tr>
<td>20-29</td>
<td>141</td>
<td>.11</td>
<td>.18</td>
</tr>
<tr>
<td>30-39</td>
<td>165</td>
<td>.15</td>
<td>.12</td>
</tr>
<tr>
<td>40-49</td>
<td>140</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>50-59</td>
<td>142</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>60-69</td>
<td>111</td>
<td>.02</td>
<td>.07</td>
</tr>
</tbody>
</table>

B. MALES

<table>
<thead>
<tr>
<th>Age at onset of illness</th>
<th>Sample Size</th>
<th>Proportion of alcoholic fathers</th>
<th>Proportion of depressive parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 20 yrs.</td>
<td>33</td>
<td>.09</td>
<td>.21</td>
</tr>
<tr>
<td>20-29</td>
<td>82</td>
<td>.08</td>
<td>.10</td>
</tr>
<tr>
<td>30-39</td>
<td>67</td>
<td>.13</td>
<td>.15</td>
</tr>
<tr>
<td>40-49</td>
<td>83</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>50-59</td>
<td>58</td>
<td>.03</td>
<td>.09</td>
</tr>
<tr>
<td>60-69</td>
<td>70</td>
<td>.04</td>
<td>.10</td>
</tr>
</tbody>
</table>
LIKELIHOOD RATIO ANALYSIS. Consider the problem of testing $H_0$ against the alternative $H_A - H_0$. In order to be specific, we will talk about the fourth set of data, namely, the proportion of depressive parents in males suffering from u.a.d. (In some ways this is the most interesting data set.) Let $\hat{p}_1; i = 1, 2, \ldots, 6$ denote the unrestricted maximum likelihood estimates of the probabilities $p_1, p_2, \ldots, p_6$ (i.e., for our data $\hat{p}_1 = .21$, $\hat{p}_2 = .10, \ldots$). The maximum likelihood estimates subject to the restrictions $H_0$ and $H_A$ can be found using any one of several algorithms in Barlow et al. (1972). Perhaps the easiest algorithm is the pool adjacent violators algorithm. Let $\hat{p}^*_i; i = 1, 2, \ldots, 6$ denote the maximum likelihood estimates subject to the restriction, $H_A$. Starting with the unsmoothed estimates $.21, .10, .15, .15, .09, .10$ the values of $\hat{p}_2$ and $\hat{p}_3$ constitute a violator since $\hat{p}_2 < \hat{p}_3$. They are both replaced by their weighted average, namely, $(n_2\hat{p}_2 + n_3\hat{p}_3)/(n_2 + n_3) = (9+10)/(87+67) = .12$ ($n_i$ is the number of probands in the $i^{th}$ group). Note that this value is obtained by "pooling" the samples from the $2^{nd}$ and $3^{rd}$ groups. We now consider the five numbers, $.21, .12, .15, .09, .10$. The violators, .12 and .15, are replaced by $(n_2\hat{p}_2 + n_3\hat{p}_3 + n_4\hat{p}_4)/(n_2 + n_3 + n_4) = .13$ and the violators .09 and .10 by $(n_5\hat{p}_5 + n_6\hat{p}_6)/(n_5 + n_6) = .09$. The resulting three numbers are decreasing in 1. Our estimates, restricted by $H_A$, are then $.21, .13, .13, .13, .09, .09$.

Let $\hat{p}_1; i = 1, 2, \ldots, 6$ denote the maximum likelihood estimates subject to $H_0$. These estimates are obtained by pooling,
the data corresponding to ages less than 40, pooling the data corresponding to ages over 40 and then pooling the resulting values if we have a reversal. For the male-depression data, the estimates are $\tilde{p}_1 = .14; i = 1, 2, 3$ and $\tilde{p}_1 = .12; i = 4, 5, 6$.

The likelihood ratio,

$$
\Lambda = \frac{\prod_{i=1}^{6} n_i \tilde{p}_1 (1-\tilde{p}_1)^{n_i(1-\tilde{p}_1)}}{\prod_{i=1}^{6} (\tilde{p}_1^*)^{n_i (1-\tilde{p}_1^*)^{n_i(1-\tilde{p}_1^*)}}}
$$

is then computed and the test rejects for large values of the statistic $T = -2 \ln \Lambda$. For this data the value of $T$ is 8.51.

The authors have recently been able to derive the appropriate limiting, null hypothesis distribution of $T$. The test based upon $T$ is not similar over the null hypothesis, $H_0$. However, one can show that if $n_i \rightarrow \infty; i = 1, 2, \ldots, 6$ in such a way that the ratios $n_i/n_j$ each converge to some strictly positive number, then

$$
\sup \lim_{n_i \rightarrow \infty} P_T = \sum_{3 \leq k \leq 6} P_1[\Lambda_1, 3]P_2[\Lambda_2, 3]P[\chi^2_{\Lambda_1 + \Lambda_2 - 2} \geq t]
$$

where $P_T$ is the probability computed under the hypothesis that $p$ is the true vector of probabilities, $\chi^2_{\Lambda}$ denotes a standard chi-square variable with $\Lambda$ degrees of freedom, and $P_1[\Lambda_1, 3]$ represents the probability that the estimates $p_1^*, p_2^*$ and $p_3^*$ assume $\Lambda_1$ distinct values. These probabilities may be computed by the following formulas (see Barlow et al., 1972):
\[ P_1(1,3) = \frac{1}{4} - \frac{1}{2n} \sin^{-1} \rho \]
\[ P_1(2,3) = \frac{1}{2} \]
\[ P_1(3,3) = \frac{1}{4} + \frac{1}{2n} \sin^{-1} \rho, \]
where
\[ \rho = -\left[ \frac{n_1n_3}{(n_1+n_2)(n_2+n_3)} \right]^{\frac{1}{2}}. \]

The values for \( P_2(L_2,3) \) are computed in like fashion. It is worth noting that even when one sample size is 3 or 4 times as large as another, the above probabilities remain close (i.e., to within .01) to the respective probabilities computed under the assumption of equal sample sizes, these being \( P(1,3) = .3333 \), \( P(2,3) = .5 \), and \( P(3,3) = .1667 \). The limiting distribution given above derives its form from the fact that it is the probability that the sum of two independent random variables exceeds the value \( t \), each variable being distributed according to Bartholomew's chi-bar-square distribution (see Barlow et al., 1972).

Table 2 contains the value of the likelihood ratio statistic and the P-values for each of the four data sets. In addition, the P-values for separately testing \( H'_0 \) vs. \( H'_A - H'_0 \) and \( H'_0 \) vs. \( H'_A - H'_0 \) are given. A combined P-value has been computed using the facts given in the introduction.

The most interesting result produced by this analysis
<table>
<thead>
<tr>
<th>TEST</th>
<th>$H_0$ vs. $H_A$</th>
<th>$H'_0$ vs $H'_A$</th>
<th>$H''_0$ vs $H''_A$</th>
<th>Combined p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-R Stat</td>
<td>9.51</td>
<td>3.54</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.015</td>
<td>0.063</td>
<td>0.073</td>
<td>0.029</td>
</tr>
<tr>
<td>Male Alcoholism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-R Stat</td>
<td>0.11</td>
<td>0.0</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.78</td>
<td>1.0</td>
<td>0.518</td>
<td>0.858</td>
</tr>
<tr>
<td>Female Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-R Stat</td>
<td>4.96</td>
<td>4.17</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.082</td>
<td>0.043</td>
<td>0.303</td>
<td>0.069</td>
</tr>
<tr>
<td>Female Alcoholism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-R Stat</td>
<td>7.94</td>
<td>0.36</td>
<td>7.60</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.019</td>
<td>0.288</td>
<td>0.007</td>
<td>0.014</td>
</tr>
</tbody>
</table>
concerns male patients and depressed parents. Neither $H_0^*$ nor $H_0^*$ can be rejected at the 5% level of significance. However, the P-value for the likelihood ratio statistic for testing $H_0$ against $H_A - H_0$ is .015. It is also interesting to note that this P-value is smaller than the "combined" P-value implying stronger evidence against $H_0$ than against $p_1 = p_2 = p_3$, $p_4 = p_5 = p_6$.

REFERENCES


**4. TITLE:** On testing two theories regarding the genetic makeup of patients suffering from unipolar affective disorder.

**7. AUTHOR(S):**
- Tim Robertson
- Giles Warrack

**9. PERFORMING ORGANIZATION NAME AND ADDRESS:**
Department of Statistics  
The University of Iowa  
Iowa City, IA 52242

**11. CONTROLLING OFFICE NAME AND ADDRESS:**
Office of Naval Research  
Statistics and Probability Program - Code 436  
Arlington, Virginia 22217

**16. DISTRIBUTION STATEMENT (of this Report):**
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

**19. KEY WORDS (Continue on reverse side if necessary and identify by block number):**
- Order restricted inference
- Hypothesis tests
- X-bar-squared distribution
- Unipolar affective disorder

**20. ABSTRACT (Continue on reverse side if necessary and identify by block number):**
In an article in the Journal of Psychiatric Research, Cadoret, Woolson and Winokur (1977) consider two theories regarding the genetic makeup of patients suffering from unipolar affective disorder. Using likelihood ratio techniques from order restricted inference we analyze their data and compare the two analyses.