

A Weighted Nonparametric Procedure for the Combination of Independent Events

JAMES A. KOZIOL

Molecular and Experimental Medicine,
Scripps Clinic and Research Foundation
La Jolla, California

HENRY C. TUCKWELL

Statistics Research Section
Australian National University

Summary

A weighted inverse normal procedure for combining independent events is considered. The procedure is utilized to combine results from six clinical studies on homovanillic acid concentrations in relation to the hypothesis that schizophrenia is the result of dopaminergic hyperactivity.

Key words: Independent events; Nonparametric combination tests; Meta analysis; Lipták procedure; Schizophrenia.

1. Introduction

Meta-analysis is an increasingly popular tool in medical research for integrating results from disparate studies into a cohesive whole. Typically, parametric procedures have been utilized for the combination of statistics summarizing "treatment" effects from different studies (e.g., COCHRAN, 1954; YUSUF et al., 1985; DERSIMONIAN and LAIRD, 1986). Nevertheless, there are circumstances under which such parametric procedures may not be of merit: distributional assumptions may not be tenable, or, more fundamentally, the various studies may be distinctly heterogeneous with regard to experimental designs, study populations, methods, or endpoints. In these circumstances, if one initially judges that, qualitatively, the studies may reasonably be combined, then omnibus nonparametric combination procedures, such as those attributed to Fisher, Tippett, and Lipták, can be used in order to screen for any effect. HEDGES and OLKIN (1985) clarify many of the qualitative issues relating to meta-analyses, and review parametric and nonparametric combination test procedures.

The omnibus nonparametric combination procedures weight the contributions of the individual experiments identically. However, BERGER and DELAMPADY

(1987) and BERGER and SELKE (1987) have argued that significance levels can depend critically on sample sizes. Hence, the purpose of this note is to suggest that a weighted Lipták procedure may be a valuable nonparametric combination procedure, with the weights corresponding to the sensitivities of the different studies. This statistic is reviewed briefly in Section 2, and optimal choice of weights is considered; it is then used to combine results of disparate studies of the dopamine hypothesis of schizophrenia in Section 3.

2. The Combination Problem

Suppose that n experiments have been carried out independently to detect a certain effect, the magnitude of which may be measured by a different parameter θ_i in each experiment. In the i -th experiment, suppose that the hypothesis $H_{0i}: \theta_i = 0$ is to be tested against the alternative $H_{1i}: \theta_i > 0$ on the basis of the real-valued test statistic T_i , where large values of T_i lead to rejection of H_{0i} . When t_i is the observed value of T_i , the attained significance level in the i -th experiment is given by

$$P_i = Pr(T_i \geq t_i | H_{0i}).$$

We assume that T_i has a continuous distribution; then, under H_{0i} , P_i is uniformly distributed on the interval $(0, 1)$. Small values of P_i lead to rejection of H_{0i} .

We consider the combination problem of determining whether the effect is present in at least one of the experiments. That is, we wish to test the combined null hypothesis

$$H_0 = \bigcap_i H_{0i}: \theta_1 = \dots = \theta_n = 0$$

versus the alternative

$$H_1 = \bigcup_i H_{1i}: \text{each } \theta_i \geq 0 \text{ with at least one } \theta_i > 0.$$

A commonly used nonparametric procedure for combining the independent significance levels P_1, \dots, P_n is LIPTÁK'S (1958) inverse normal procedure: one would reject H_0 at level α if

$$L = - \sum_{i=1}^n \Phi^{-1}(P_i) \geq n^{\frac{1}{2}} \Phi^{-1}(1 - \alpha),$$

where Φ is the standard normal cumulative distribution function.

A generalization of Lipták's procedure is to incorporate weights $w_i \geq 0$ into L , that is,

$$L_w = \frac{-\sum_{i=1}^n w_i \Phi^{-1}(P_i)}{\left(\sum_{i=1}^n w_i^2\right)^{\frac{1}{2}}};$$

one would reject H_0 for $L_w \geq \Phi^{-1}(1 - \alpha)$. The weights w_i should be chosen to reflect the efficiencies of the test statistics used in the individual experiments. In particular, a useful choice of w_i is the inverse of the standard deviation σ_i of T_i . This choice of weights is readily motivated: for example, if the statistics T_i are normally distributed with means θ_i and known variances σ_i^2 , and the true effects θ_i are assumed equal to a common value θ , then the best linear unbiased estimator of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^n T_i/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}.$$

Since $T_i/\sigma_i = -\Phi^{-1}(P_i)$, it follows that

$$L_w = \frac{-\sum_{i=1}^n \Phi^{-1}(P_i)/\sigma_i}{\left[\sum_{i=1}^n 1/\sigma_i^2\right]^{\frac{1}{2}}}$$

is the best test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ on the basis of $\hat{\theta}$.

COCHRAN (1954) and DERSIMONIAN and LAIRD (1986) have proposed $\hat{\theta}$ as an estimator in the fixed effects model described above. Alternatively, under a random effects model, in which the θ_i are normally distributed about θ with variance τ^2 , they proposed an estimator

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i T_i}{\sum_{i=1}^n w_i},$$

of θ , with weights

$$w_i = (\sigma_i^2 + \tau^2)^{-1}.$$

Again, with the choice of weights the reciprocals of the standard deviations

$$(\sigma_i^2 + \tau^2)^{-\frac{1}{2}},$$

it is clear that L_w would appropriately assess $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ on the basis of $\hat{\theta}$. That is, under these models, L_w optimally assesses the departure of $\hat{\theta}$ from zero.

To investigate further the statistical properties of the weighted inverse normal procedure, we introduce a structural model for the results of the n independent studies. Suppose in each study a control group C is compared with an experimental group E on the basis of observations X_{ij} , $j = 1, \dots, n_{Ci}$ for the i -th control group, $i = 1, \dots, n$ and Y_{ij} , $j = 1, \dots, n_{Ei}$, $i = 1, \dots, n$, for the i -th experimental group, where n_{Ci} and n_{Ei} are the sample sizes in the control and experimental groups, respectively. Suppose further that $E(X_{ij}) = \mu_{Ci}$, $E(Y_{ij}) = \mu_{Ei}$, and $\text{Var}(X_{ij}) = \text{Var}(Y_{ij}) = \sigma_i^2$. Define the effect size for the i -th study as

$$\delta_i = (\mu_{Ei} - \mu_{Ci}) / \sigma_i.$$

Let

$$T_i = (\bar{Y}_i - \bar{X}_i) / s_i \sqrt{\frac{1}{n_{Ei}} + \frac{1}{n_{Ci}}},$$

the two sample t -statistic with $n_{Ei} + n_{Ci} - 2$ degrees of freedom for assessing the null hypothesis that $\mu_{Ei} = \mu_{Ci}$; here, s_i^2 is the usual pooled estimate of σ_i^2 . In terms of the combination problem, one might take $\theta_i = \delta_i$, or equivalently $\theta_i = \lambda_i$, where

$$\lambda_i = \delta_i \sqrt{\frac{n_{Ci} n_{Ei}}{n_{Ci} + n_{Ei}}}.$$

Let P_i be the one-tail null p -value corresponding to T_i . Then $Z_i = \Phi^{-1}(1 - P_i)$ is approximately normally distributed with mean λ_i with variance 1 if the sample sizes are relatively large (ROSENTHAL and RUBIN, 1979). Although one can proceed to test for homogeneity among the magnitudes of the effects (and estimate a common effect size, on the basis of the inverse normal transformation (ROSENTHAL and RUBIN, 1982)) we focus here on the use of weights in Lipták's combination procedure. That is,

$$L_w = \sum_{i=1}^n w_i Z_i / \left(\sum_{i=1}^n w_i^2 \right)^{\frac{1}{2}}$$

is normally distributed with mean

$$\sum w_i \lambda_i / \left(\sum w_i^2 \right)^{\frac{1}{2}}$$

and variance 1; note that the mean is maximized if $w_i = \lambda_i / \sqrt{\sum \lambda_i^2}$. The relative efficiency of the unweighted (i.e., $w_i = 1$) Lipták procedure compared to the optimal weighted procedure is therefore

$$\text{eff}(L, L_w) = \frac{\left(\sum_{i=1}^n \lambda_i \right)^2 / n}{\sum_{i=1}^n \lambda_i^2} \leq 1.$$

If the effect sizes δ_i are all equal, and there is a common experimental group sample size and a common control group size, then the λ_i are all equal and the unweighted Lipták procedure is fully efficient. If the effect sizes are equal but the sample sizes are unequal, one should use weights

$$w_i = \sqrt{\frac{n_{Ci} n_{Ei}}{n_{Ci} + n_{Ei}}}$$

for optimality. If the effect sizes and sample sizes are unequal, one can always achieve full efficiency by optimal weighting, but the magnitude of this gain relative to any other weighting scheme needs to be examined on a case-by-case basis.

3. The Dopamine Hypothesis in Schizophrenia

The interaction of antipsychotic drugs with monoamine neuronal receptors suggests a role for monoamine transmission mechanisms in the pathophysiology of schizophrenia. In particular, a biochemical dysfunction in schizophrenia is posited by the dopamine (DA) hypothesis, which implicates an overactivity of certain central nervous system (CNS) dopaminergic neurons, or alternatively, an increased sensitivity of the corresponding dopamine receptors in the etiology of this disorder. (See MELTZER and STAHL, 1976, and HARACZ, 1982, for reviews of the DA hypothesis.) Typically, levels of neurotransmitter precursors and metabolites have been measured in the cerebrospinal fluid (CSF), plasma, or urine in order to assess central neurochemical function. The predominant extraneuronal DA metabolite is homovanillic acid (HVA): from BOWERS (1972), HVA levels represent processes of formation of the metabolite, that is, DA turnover, thus are indicative of the presynaptic activity of dopaminergic neurons in the CNS. See AMIN et al. (1992) for further discussion of this claim.

HERITCH (1990) has qualitatively summarized 19 clinical studies that had examined DA metabolite levels in neuroleptic-free schizophrenic patients. In this regard, given the variability in physiologic parameters found in schizophrenics and normal individuals, well-defined cohorts of patients should be studied and compared with control groups well-matched to the patients on such factors as age and sex. On this basis, we screened the studies reviewed by Heritch to focus on six of the studies in which comparisons were made between schizophrenics and normal controls who were well-matched in terms of age and gender. We extracted information on HVA levels in the patient and control groups from each of these studies: under the DA hypothesis, one would expect HVA levels to be greater among schizophrenics than among normal individuals. These summary data are presented in Table 1. Our aim here is to examine the validity of the DA hypothesis quantitatively with these data, complementary to Heritch's qualitative analysis.

Table 1
Studies of Dopamine Metabolite Levels in Drug-Free Schizophrenic Patients and Normal Controls

Study	Outcome	Effect Size	n_E	n_C	P
Berger et al., 1980	CSF HVA	.029	8	20	.473
Nybäck et al., 1983	CSF HVA	.083	26	43	.370
Gerner et al., 1984	CSF HVA	.159	20	37	.285
Lindström, 1985	CSF HVA	-.607	40	21	.986
Karoum et al., 1987	24-hr urinary excretion of DA + metabolites	-.757	11	16	.968
Davidson and Davis, 1988	plasma HVA (average over 12 hrs)	-.986	14	14	.993

CSF HVA was typically measured in ng/ml or pmol/ml, 24-hr urine values were in units of $\mu\text{mol/day}$, and plasma HVA was measured in ng/ml. The effect size for the i -th study is the standardized mean difference

$$d = \frac{\bar{X}_E - \bar{X}_C}{s},$$

where \bar{X}_E and \bar{X}_C are the sample means of the outcome variable for the schizophrenic and control groups, with sample sizes n_E and n_C , respectively, in the i -th study, and s is the pooled sample standard deviation

$$s = \sqrt{\frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2}}.$$

The p -value is the one-sided exceedance probability for the statistic d upon reference to a central t sampling distribution; that is,

$$P = \Pr\left(T \geq \sqrt{\frac{n_E n_C}{n_E + n_C}} d\right), \quad T \sim t_{n_E + n_C - 2}$$

Since the study outcomes are heterogeneous, Lipták's weighted combination procedure will be used to assess the overall evidence of the DA hypothesis. One might weight the individual studies according to some measure of their quality; see DERSKY et al. (1992) for further discussion. We judge the studies to be of comparable quality, and expect the effect sizes of dopamine metabolite levels to be comparable. We therefore weight the contribution of the i -th study by

$$w = \sqrt{\frac{n_E n_C}{n_E + n_C}},$$

the inverse of the standard deviation of the effect size for that study, as discussed in the previous section. Using these weights, we calculate

$$L_w = - \sum_{i=1}^6 w_i \Phi^{-1}(P_i) / \left(\sum_{i=1}^6 w_i^2 \right)^{\frac{1}{2}} = -1.998,$$

yielding a one-sided p -value of 0.977. This result does not support the claim that homovanillic acid levels are raised in schizophrenics and hence indirectly does not lend support to the dopamine hypothesis. To the contrary, the present finding suggests that DA turnover may actually be reduced in schizophrenia. One might postulate that in schizophrenics a reduction in DA release can induce postsynaptic receptor supersensitivity, which in turn could give rise to hyperactivity and a concomitant acute phase of the disorder at times of increase in DA release rate (MACKAY, 1980).

References

- AMIN, M., DAVIDSON, M. and DAVIS, K. L., 1992: Homovanillic acid measurement in clinical research: A review of methodology. *Schizophr. Bull.* **18**, 123–141.
- BERGER, J. O. and DELAMPADY, M., 1987: Testing precise hypotheses (with discussion). *Statist. Sci.* **2**, 317–352.
- BERGER, J. O. and SELLEKE, T., 1987: Testing a point null hypothesis: The irreconcilability of P values and evidence (with discussion). *J. Amer. Statist. Assoc.* **82**, 112–139.
- BERGER, P. A., FAULL, K. F., KILKOWSKI, J., et al., 1980: CSF metabolites in depression and schizophrenia. *Amer. J. Psychiatry* **137**, 174–180.
- BOWERS, M. B., 1972: Clinical measurements of central dopamine and 5-hydroxytryptamine metabolism: Reliability and interpretation of cerebrospinal fluid acid monoamine metabolite measures. *Neuropharmacology* **11**, 101–111.
- COCHRAN, W. G., 1954: The combination of estimates from different experiments. *Biometrics* **10**, 101–129.
- DAVIDSON, M. and DAVIS, K. L., 1988: A comparison of plasma homovanillic acid concentrations in schizophrenic patients and normal controls. *Arch. Gen. Psychiatry* **45**, 561–563.
- DETSKY, A. S., NAYLOR, C. D., O'ROURKE, K., et al., 1992: Incorporating variations in the quality of individual randomized trials into meta-analysis. *J. Clin. Epidemiol.* **45**, 255–262.
- GERNER, R. H., FAIRBANKS, L., ANDERSON, G. M., et al., 1984: CSF neurochemistry in depressed, manic, and schizophrenic patients compared with that in normal controls. *Amer. J. Psychiatry* **141**, 1533–1540.
- HARACZ, J. L., 1982: The dopamine hypothesis: An overview of studies with schizophrenic patients. *Schizophr. Bull.* **8**, 438–469.
- HEDGES, L. V. and OLKIN, I., 1985: *Statistical Methods for Meta-Analysis*. Academic Press, Orlando.
- HERITCH, A. J., 1990: Evidence for reduced and dysregulated turnover of dopamine in schizophrenia. *Schizophr. Bull.* **16**, 605–615.
- KAROUM, F., KARSON, C. N., BIGELOW, L. B., et al., 1987: Preliminary evidence of reduced combined output of dopamine and its metabolites in chronic schizophrenia. *Arch. Gen. Psychiatry* **44**, 604–607.

- LINDSTRÖM, L. H., 1985: Low HVA and normal 5HIAA CSF levels in drug-free schizophrenic patients compared to healthy volunteers: Correlations to symptomatology and family history. *Psychiatry Res.* **14**, 265–273.
- LIPTÁK, T., 1958: On the combination of independent events. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **3**, 171–197.
- MACKEY, A. V. P., 1980: Positive and negative schizophrenic symptoms and the role of dopamine. *Brit. J. Psychiatry* **137**, 379–383.
- MELTZER, H. Y. and STAHL, S. M., 1976: The dopamine hypothesis of schizophrenia: A review. *Schizophr. Bull.* **2**, 19–76.
- NYBÄCK, H., BERGGEN, B.-M., HINDMARSH, T., et al., 1983: Cerebroventricular size and cerebrospinal fluid monoamine metabolites in schizophrenic patients and healthy volunteers. *Psychiatry Res.* **9**, 301–308.
- ROSENTHAL, R. and RUBIN, D. B., 1979: Comparing significance levels of independent studies. *Psychol. Bull.* **86**, 1165–1168.
- ROSENTHAL, R. and RUBIN, D. B., 1982: Comparing effect sizes of independent studies. *Psychol. Bull.* **92**, 500–504.
- YUSUF, S., PETO, R., LEWIS, J., COLLINS, R., and SLEIGHT, P., 1985: Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Prog. Cardiovasc. Dis.* **27**, 335–371.

Received, Sept. 1993

Accepted, Nov. 1993

JAMES A. KOZIOL, Ph. D.
Scripps Clinic and Research Foundation
10666 North Torrey Pines Road
La Jolla, CA 92037
USA