demonstrated by this paper) for which we have very little in the way of convincing explanations?

DAVID CURTIS
St Mary's Hospital Medical School
Academic Department of Psychiatry
St Mary's Hospital
Praed Street
London W2 1NY

AUTHORS' REPLY: The first question raised by Dr Curtis is concerned with the interpretation of the parameter estimates of the log-linear model. We did not make explicit in our paper that natural logarithm (base e), rather than common logarithm (base 10), was used. On the other hand, the use of natural logarithm in log-linear models is almost universal, and this would be consistent with our statement that "The estimate for influenza corresponds to an increase by a factor of $e^{0.000014}$ in the number of schizophrenic births for every death attributed to influenza two to three months previously". What this means is that, suppose a certain month would have 'experienced' 100 schizophrenic births without the influence of influenza, then a single influenza death 2–3 months earlier would have increased the number of schizophrenic births, on average, to $100 \times e^{0.000014}$. The value of $e^{0.000014}$ is approximately 1.000014, so that the number of schizophrenic births would be increased, on average, to 100.0014. For 1000 influenza deaths, this becomes approximately 101.4, which corresponds to a 1.4% increase.

The second question on multiple testing touches on some fundamental and controversial issues in statistics. The usual advice in such situations is to make an adjustment, such as the Bonferroni correction, to each separate comparison, or to perform an 'omnibus' test which covers all the comparisons, and consider individual comparisons only if the overall test is significant. However, others believe that no adjustment for multiple testing is necessary. For example, Rothman (1986) states "Since no problem calling for any adjustment seems to exist unless the positive results from a large number of comparisons are reported without any information about the total number of comparisons, and since even then it appears that adjustments in the results only make them more difficult to interpret, the best course of action for the epidemiologist to take when making multiple comparisons is to ignore advice to make such adjustments in reported results."

The rationale for the Bonferroni correction is that, in making m independent tests, the probability of at least one significant result (at the conventional level of $P < 0.05$) when in fact all the null hypotheses are true is not 0.05, but $1-(0.95)^m$, or approximately $0.05\,m$. Thus, to limit the probability of erroneously rejecting *any* null hypothesis to less than 0.05, the P-values of the individual tests are multiplied by m. However, by making the test more stringent, this procedure also increases the probability of not rejecting a false null hypothesis.

In a complex data set, should the evidence for any particular hypothesis depend on how many other hypotheses are also examined? If the answer is yes, then one can destroy the evidence for almost any hypothesis by increasingly thorough analysis of the data. Conversely, the investigator who analyses and publishes his hypotheses one-by-one will run into no such problem. Clearly this is absurd. We believe that it is not the number of tests which is crucial, but whether there are compelling reasons, based on existing knowledge, for making the predictions. A hypothesis supported by a priori knowledge should be considered in its own right, regardless of how many other tests are performed. However, when there is no previous support for a hypothesis, any positive evidence should be considered tentative, even if only one test is performed. There is a subjective element in deciding whether there is an a priori basis for hypothesis, but this can be incorporated into the modern Bayesian framework of scientific inference (Howson & Urbach, 1988).

Having said this, it is often true, when an investigator makes a large number of tests, that there is no a priori basis for any of the hypotheses. Indeed, under some circumstances, there is an inverse relationship between the number of tests and the a priori probability of each hypothesis. For example, consider two random loci which are being tested for allelic association with a disease. There should be no difference in the prior probabilities that each locus is in linkage equilibrium with a disease locus. However, a chance association may be more likely with the more polymorphic locus, which contains more alleles for comparison. An adjustment for multiple tests will then reduce the risk of becoming over-confident about any 'significant' results.

The apparent need for an adjustment for multiple tests is a reflection of the current emphasis placed on the P-value, of a single study, in determining the truth of a hypothesis. Unless the effect or sample size is enormous, this is rarely possible when there is substantial random variation. The truth of a hypothesis must then be assessed by repeated, independent studies; an assessment that should consider all the studies, positive or negative, but that need not be influenced by whether other hypotheses are also addressed in these studies. Chance findings

https://doi.org/10.1192/S0007125000130466 Published online by Cambridge University Press

are unlikely to be replicated across independent studies.

In making 10 tests of the relationship between schizophrenic births and influenza deaths, we aimed to cover the entire period of gestation plus the month of birth. We certainly did not regard the tests as equally likely to show a significant result, since neuropathological evidence points to a disturbance of neurodevelopment in the second half of gestation (Roberts, 1991), and several previous studies have found a relationship between second or early third trimester exposure to influenza epidemics and schizophrenia (Mednick *et al*, 1988; Barr *et al*, 1990; O'Callaghan *et al*, 1991). Bearing in mind the existing evidence, the *P*-values of 0.02 for the relationships between schizophrenic births and influenza deaths 2 and 3 months previously should be considered as further evidence in support of the maternal viral infection hypothesis. Indeed, the lack of a significant effect of influenza in the first few months of gestation is exactly what we could have predicted from the existing evidence. Nonetheless, we realise that our results alone are not conclusive, hence our statement that our report "adds to the current evidence for an association between prenatal exposure to influenza epidemics and the later development of schizophrenia".

BARR, C. E., MEDNICK, S. A. & MUNK-JORGENSEN, P. (1990) Exposure to influenza epidemics during gestation and adult schizophrenia: a 40-year study. *Archives of General Psychiatry*, **47**, 869–874.

HOWSON, C. & URBACH, P. (1988) *Scientific Reasoning: the Bayesian Approach*, Illinois: Open Court Publishing Company.

MEDNICK, S. A., MACHON, R., HUTTENEN, M. O., *et al* (1988) Adult schizophrenia following prenatal exposure to an influenza epidemic. *Archives of General Psychiatry*, **45**, 189–192.

O'CALLAGHAN, E., SHAM, P. C., TAKEI, N., *et al* (1991) Schizophrenia after prenatal exposure to 1957 A2 influenza epidemic. *Lancet*, **337**, 1248–1250.

ROBERTS, G. (1991) Schizophrenia: a neuropathological perspective. *British Journal of Psychiatry*, **158**, 8–17.

ROTHMAN, K. J. (1986) *Modern Epidemiology*. Boston: Little, Brown and Company.

P. C. SHAM
N. TAKEI
R. M. MURRAY
Department of Psychological Medicine
Institute of Psychiatry & King's College Hospital
De Crespigny Park
London SE5 8AF

E. O'CALLAGHAN
Cluain Mhuire Family Centre
Blackrock
Co. Dublin
Republic of Ireland

Ventricle–brain ratio in schizophrenia

SIR: I read with interest the paper by van Horn & McManus (*Journal*, May 1992, **160**, 687–697). The most interesting finding of the study is the negligible difference in ventricle–brain ratio (VBR) between patients with schizophrenia and control subjects. Testing the validity and reliability of the conducted meta-analysis by attempting to answer the ten questions put by Wilson & Henry (1992*a*,*b*) indicates that the results of this meta-analysis are questionable. The authors concluded, erroneously, that the differences in VBR between patients and controls were smaller in studies where more stringent criteria for schizophrenia were employed, although they did not comment that such differences may be due to more than errors of measurements. The authors pooled data for meta-analysis regardless of the methods of measurements used. To be more specific, the authors came to a conclusion that the largest difference in VBR between patients and controls was found in studies that used DSM–III criteria. There were 19 studies which found significant differences in VBR between patients and controls, and patients were diagnosed along the lines of DSM–III; in 11 of these, computerised measurement of VBR was used, and in 8 studies hand planimetry was used. If schizophrenic patients have significantly larger ventricles than normal controls one would expect with a high probability that the grand mean of VBR of patients included in the 11 and 8 studies, would be higher than the grand mean of VBRs of the respective controls. It happened that the ages and standard deviations of VBR of patients and controls in the two groups of samples were comparable. We reanalysed the presented data. The first two-sample analysis of the difference between the grand means of patients and controls (11 studies, computerised method of measurements) revealed no significant difference ($t = 2.03$, $P > 0.056$; 95% CI for difference in means -0.049–3.660 and difference between means 1.8); nonparametric two-sample analysis revealed also no significant difference in the median or in ranks. The second two-sample analysis revealed a significant difference between the grand means of patients and controls (8 studies, hand planimetry as a method of measurements; $t = 3.93$, $P < 0.01$, 95% CI for difference in means 1.10–3.76, difference between means 2.43). We plotted the means of VBRs of patients and controls in the two groups and found that the distribution of means of VBR of patients and controls where a computerised method of measurement was used was identical, but the plot of means of VBRs of patients and controls in the second group reveals a different distribution of means. We postulate that the difference in VBR between

https://doi.org/10.1192/S0007125000130466 Published online by Cambridge University Press