Brit. J. prev. soc. Med. (1971), 25, 174-176

ON EDWARDS' CRITERION OF SEASONALITY AND A NON-PARAMETRIC ALTERNATIVE

DAVID HEWITT, JEAN MILNER, ADELE CSIMA, AND ANDREW PAKULA

Department of Epidemiology and Biometrics, School of Hygiene, University of Toronto

There are a few conditions—hay fever will serve as an example—for which it has always been possible to assert, without fear of contradiction, that their pattern of occurrence was a seasonal one. However, when it is not subjectively clear that a seasonal pattern exists, considerable difficulty can be experienced. Applied to data in the form of monthly frequencies of new cases, the common χ^2 test with 11 degrees of freedom will not be sensitive to seasonal fluctuations of moderate amplitude unless the sample size is very large. Moreover, when this test does detect a significant amount of variation between months it will still be left to the investigator to judge whether the variation is of a seasonal character. When a prior hypothesis is available it is, of course, easy to apply a more efficient test; for example, if it has previously been suggested that a disease tends to occur more frequently from April to September then one could test whether the proportion of cases in a fresh sample that occurs within this period exceeds 0·501 (=183/365·25). In the situation more usually encountered there is no prior hypothesis, and the same limited sample of cases must be used both to support a test of the null hypothesis and to specify a likely type of seasonality.

The first convenient method of dealing with this situation was devised by Edwards (1958, 1961), who proposed restricting consideration to patterns that could be tolerably well described by a simple harmonic curve. His method of analysis yields estimates of two parameters of this curve and simultaneously provides a test criterion whose distribution, on the null hypothesis, approximates to that of χ^2 with 2 degrees of freedom. The method has recently been criticized (Wehrung and Hay, 1970) on the ground that it may be sensitive to some kinds of cyclic variation that are clearly not of simple harmonic form. So far from being a fault, this seems to us to be a virtue of the method, and one that tends to weaken the case for introducing a non-parametric test like the one proposed below. That Edwards' criterion may also be sensitive to occasional, non-recurrent outbreaks is a risk acknowledged by its author, and one less likely to afflict a non-parametric test. Both types of test are liable to give a false positive result if used on data that are subject to any strong secular influence, since this will suggest the presence of a 'seasonal' peak in September-October or in March-April, depending on whether the trend is rising or falling.

We were first led to reconsider the performance of Edwards' criterion after having used it, perhaps injudiciously, as a screening device to pick out, from among a large number of congenital cardiac malformations, those particular defects in which the influence of seasonally varying factors could be regarded as most plausible. A conspicuous feature of this attempted screening was that a much higher proportion of apparently significant results was declared when the sample of cases was small. It was at first hoped that smallness of a diagnostic subgroup (in a total series of over 10,000 affected children) might connote aetiological homogeneity, and hence a relatively good chance of detecting genuine seasonal variation, but prudence obviously required that we should examine the adequacy of the χ^2 approximation for the smaller sample sizes.

Edwards' original papers did not refer to the question of requisite sample size, and the note by Smith (1961), which accompanied the second paper, only indicated that it should be 'reasonable' and 'sufficiently large'. We therefore set up a Monte Carlo experiment using pseudo-Poisson variate values in sets of 12, and carried out a thousand runs for each of seven values of the 'true' mean monthly frequency, corresponding to sample sizes from 192 down to 24. A selection from the computer output is shown in Table I, where it will be seen that there was some excess of apparently significant results at all the sample sizes considered, but that the approximation was very good down to sample size 96. Even for samples as small as 48 the criterion might be considered a serviceable one, though the frequency of values beyond the upper 10% and 5% point of the χ^2 distribution was about twice, and that of values beyond the 1% point about three times what it

TABLE I

CUMULATIVE DISTRIBUTION OF EDWARDS' CRITERION COMPARED WITH THAT OF χ^2 ON 2 DEGREES OF FREEDOM (RESULTS FROM SETS OF 1,000 COMPUTER RUNS USING PSEUDO-POISSON INPUT)

Mean no. of cases per month		16	8	7	6	5	4	2
Corresponding sample size		192	96	84	72	60	48	24
Value of criterion	$P\{\chi^2_{(2)}\}$							
9·210	0·010	11	14	17	17	33	31	118
5·991	0·050	54	58	62	70	95	99	237
4·605	0·100	107	116	120	141	157	188	332
1·386	0·500	502	510	517	522	533	557	702

should have been. Unfortunately, several of the diagnostic groups to be reviewed had less than 50 cases. We therefore felt the need for some alternative statistical criterion of seasonality for which the type I error would be, as far as possible, independent of sample size.

That Edwards had also envisaged the possible use of a non-parametric test is shown by his allusion to the cyclic sequence of months with frequencies above and below the median. His paper includes an expression for the chance probability of encountering six successive months (e.g., November to April) with values on the same side of the median* and indicates that this could afford the basis of a 'simple but inefficient test'. We judged that a somewhat more efficient test could be achieved by using all the ranking information rather than a simple dichotomy and decided to try a criterion based on the sum of the ranks of r successive months. On the ground that the chance probability of obtaining the largest possible rank sum is smallest when $r = n/2$ we chose to use the rank sum for six successive months. Some characteristics of this six-month rank-sum criterion are shown in Table II, where it is assumed that the monthly frequencies (or preferably the monthly incidence rates) have been ranked from 12 (highest) down to 1 (lowest).

It is again appropriate to distinguish the situation in which a prior hypothesis, specifying the six-month period of higher expected incidence, is available for testing, from the situation in which the likely nature of any seasonality has to be inferred from the data. For the former situation the exact distribution of the rank-sum criterion is easily derived by enumeration of the 924 possible combinations of ranks that can occur in a pre-assigned six-month segment of the year. The rank-sum is distributed symmetrically over the range 21 to 57 (see first column of Table II). For purposes of significance testing one would refer to the upper tail of the cumulative distribution (see second column of Table II). A value of the test criterion equal or greater than 50 would be regarded as significant at the conventional 5% level.

When no prior hypothesis exists it is necessary to

*However, the actual expression given is incorrect. There are altogether (n–l)! different cyclical permutations of n objects. If r of the objects belong to one class, the proportion of permutations in which all members of this class occur in an unbroken sequence is r! (n–r)! /(n–l)!

TABLE II

CHARACTERISTICS OF THE SUGGESTED RANK-SUM CRITERION OF SEASONALITY

	Null Distributions				Test Results	
	Rank-sum for a Pre-assigned Segment of 6 Months		Largest Rank-sum for any 6-month Segment		Largest Rank-sum for Any 6-month Segment based on samples of 50 or more children with congenital heart disease for which $P\{\chi^2\}$ by Edwards' criterion was:	
	Exact Distribution	Cumulative Probability	Distribution from 5,000 'Monte Carlo' Trials	Estimated Cumulative Probability	<0·10	>0·10
57	1	0·0011	67	0·0134*	2	
56	1	0·0022	57	0·0248	1	
55	2	0·0043	108	0·0464	1	
54	3	0·0076	151	0·0766	1	
53	5	0·0130	247	0·1260	1	1
52	7	0·0206	327	0·1914	1	
51	11	0·0325	497	0·2908		
50	13	0·0465	459	0·3826		6
49	18	0·0660	566	0·4958		4
48	22	0·0898	564	0·6086		5
47	28	0·1201	586	0·7258		2
46	32	0·1548	526	0·8310		2
45	39	0·1970	414	0·9138		4
44	42	0·2424	238	0·9614		1
43	48	0·2944	145	0·9904		1
42	51	0·3496	41	0·9986		
41	55	0·4091	7	1·0000		
40	55	0·4686				
39 or less	491	1·0000				
Total	924		5000		7	26

*Exact value: 1/77=0·0130

locate the six-month segment which yields the highest value of the rank-sum. The null distribution of this *maximum* rank-sum is not easy to formulate or to enumerate, even on a high-speed computer. We have, therefore, again resorted to Monte Carlo methods, this time generating and testing 5,000 random permutations of the numbers 1 to 12, with the results summarized in the third and fourth columns of Table II. It is a little disappointing that only the three highest values of the maximum rank-sum criterion (representing four different combinations of ranks) warrant the verdict of significance at the conventional 5% level.

The application of the proposed non-parametric test may be illustrated by data derived from the large registry of cardiac malformations, of all types, kept at the Hospital for Sick Children, Toronto, in which the patient's date of birth is routinely recorded. By relating these dates to the month-of-birth distribution in the population from which the patients are drawn, estimates of relative incidence were calculated for each month and found to be in the following rank order:

Jan. 5; Feb. 12; March 7; Apr. 4; May 6; June 2:
July 3; Aug. 9; Sept. 1; Oct. 11; Nov. 10; Dec. 8.

The maximum value of the rank-sum is that obtained for the six-month period running from October to March, where $11+10+8+5+12+7 = 53$. As Table II shows, approximately one sample in eight would, on a chance basis, be expected to yield a value as large as 53. However, there is another perspective in which this particular result might be viewed. Maternal rubella is a recognized cause of one type of cardiac malformation*, and the seasonal distribution of rubella in Ontario is such that any teratogenic effect exerted at about the third week of gestation would be expected to express itself in an excess of affected infants among those born in the period October to March (Bell, 1969). Hence, if the analysis had been intended for the purpose of detecting a hypothetical influence of rubella on cardiac malformations in general, this result might be quoted as having a degree of statistical significance measured by the value of $P = 0{\cdot}013$.

The practical utility of the rank-sum criterion will be for others to judge. On the basis of our experience to date we find it appropriate, for the time being, to use the non-parametric and the well-established parametric method in parallel, but to discount results from the latter when the sample size falls below 50. We set no lower limit on the sample size for the rank-sum criterion, except as entailed by the practice of requiring that at least six of the 12 months must have non-zero frequencies. When calculating monthly incidence rates we prevent any tying of ranks by adding 0·001 of a case to each numerator and using a three-digit denominator.

Near equivalence of the two methods, when applied to samples of over 50 cases, is suggested by the comparison illustrated in the final columns of Table II. It is there shown that the rank-sum criterion achieved quite good separation between samples in which the evidence for 'harmonic' variation was or was not suggestive at the 10% level of significance.

*Cases of patent ductus arteriosus with a history of maternal rubella were excluded from the case series for the purpose of the analysis described.

We are greatly indebted to Dr. Vera Rose for bringing this problem to our attention and for permission to make use of data from the Department of Cardiology of the Hospital for Sick Children, Toronto. This work has been supported by funds from the National Health Grants Programme (Grants Nos. N–605–7–434 and N–605–7–636).

REFERENCES

Bell, J. S. (1969). Personal communication.

Edwards, J. H. (1958). Congenital malformations of the central nervous system in Scotland. *Brit. J. prev. soc. Med.*, **12**, 115.

—— (1961). The recognition and estimation of cyclic trends. *Ann. hum. Genet.*, **25**, 83.

Smith, C. A. B. (1961). Note on the error variance. *Ann. hum. Genet.*, **25**, 86.

Wehrung, D. A., and Hay, S. (1970). A study of seasonal incidence of congenital malformations in the United States. *Brit. J. prev. soc. Med.*, **24**, 24.