Variola minor in Braganca Paulista County, 1956: household aggregation of the disease and the influence of household size on the attack rate

By JUAN J. ANGULO*

Instituto Adolfo Lutz, São Paulo, S.P., Brazil

AND STEPHEN D. WALTER

Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut, USA

SUMMARY

Household aggregation of cases, one possible characteristic of person-to-person transmitted disease, was formally tested in one epidemic of variola minor by using a pair statistic. A significant result was found for all households as well as for households grouped by the type of environment, or by the phase of the epidemic growth in time. Secondary attack rates, when related to household size (number of susceptibles) showed only a marginal trend in rural households but no trend in urban or semi-rural households.

INTRODUCTION

The appearance of multiple cases of disease in the same house has been intuitively related to contagion. This is on the basis that transmission of the disease is more likely between members of the same household than between members of different households. A statistical demonstration is thus required for objectively demonstrating the clustering. The theory of disease clustering in houses was developed early in this century (Pearson, 1912; Troup & Maynard, 1912). In spite of the voluminous literature on variola, a statistical study of clustering of variola cases in households has not been attempted although the phenomenon has recently been analysed in schools (Klauber & Angulo, 1974). As a matter of fact, clustering of variola cases in households has rarely been noted (Dixon, 1948) until Angulo *et al.* showed that in certain communities the real setting of a variola epidemic was the household (Angulo, Rodrigues-de-Silva & Rabello, 1967, 1968), a finding promptly confirmed by various workers (Henderson & Yepke, 1969; Heiner, Fatima & McCrumb, 1971; Arnt & Morris, 1972). The present paper reports the results of significance tests for clustering using Walter's pair statistic (Walter, 1974), as applied to the numbers of cases of variola minor appearing in the households affected by the epidemic studied in the present paper.

<div align="center">MATERIALS AND METHODS</div>

Basic data

A general description of the study epidemic and of the methods employed during its field study has been presented earlier (Angulo, 1976). All households for which cases of variola were notified were surveyed. Full cooperation from the public and authorities permitted an unusual degree of completeness of notifications to be reached. Cases occurring in each household were recorded and the characteristics of all persons from each household, whether with or without variola, were determined. However, limitations in transport and time meant that not all surveyed households could be kept under surveillance. These exceptions were neglected in the present study together with rare households whose data were incomplete for the reasons pointed out. With the resources available, it was not possible to enumerate the several thousand households where no case occurred; also the cadastral city map was of limited usefulness because of the large number of business or empty dwellings, even in the most densely populated subdivision of the city. No cadastral map of the rural districts was available.

Clustering of cases in households was studied on the whole data set, as well as subdivided according to certain factors, which on intuitive grounds might have been thought to influence the degree of clustering. These factors were characteristics of the environment of the household and the phases or stages of the epidemic development with time.

The geographical position of the affected household dwellings was recorded. Three distinct types of environment which correspond to different hygienic and social standards of living were distinguished. Rural households were those living in farms, outside the official perimeter of the city capital of the County. Urban households were those living in the city core, a centrally located subdivision which had the highest density of dwellings, a water supply system and the highest hygienic standards. Most commercial and social activities concentrated in the city core. Semi-rural households lived at the periphery of the city core, within the official perimeter of the city. Here, distances between dwellings were greater, hygienic conditions generally poor and there was a limited community life.

The following phases could be distinguished in the development of the epidemic with time. Phase I went from onset of the epidemic (on 10 November 1955) to opening of the schools (on 3 March 1956) for the 1956 academic year. Phase II was the first 'semester' of school activities (from 3 March to 30 June). Phase III corresponded to closure of the schools for the regular Winter vacation (1–31 July). Phase IV corresponded to part of the second 'semester' of school activities and went from 1 August to 1 October 1956. On the latter date, the last case of the epidemic had appeared.

Analysis

Walter's pair statistic (1974) was used for formally testing household aggregation of variola cases. The test consists of a comparison of the observed number of distinct pairs of cases within the same households with the expected number

of such pairs under the null hypothesis that the cases are randomly distributed among all members of the population under study. Clustering is indicated when the observed number of pairs is significantly greater than its expectation, the significance being given by the value of a normal distribution z-score based on the estimated mean and variance of the number of pairs. If the number of cases of disease in a house is denoted by r, and the total number of members of a household by m, then to derive the mean and variance of the pair statistic, one needs to calculate Σr, Σr^2, Σm, Σm^2, and Σm^3, all summations taken over all households in the data being examined in a particular test. If the total number of houses is denoted by S, and defining $R = \Sigma r$, $M = \Sigma m$, then the expected number of pairs T is given by

$$E(T) = \frac{(R-S)[(R-S-1)\Sigma m^2 - 3MR - MS + M + 2RS + 2M^2]}{2(M-S)(M-S-1)}. \qquad (1)$$

Similarly the variance of T is given by

$$V(T) = (R-S)(R-S-1)(M-R)[2(M-S)(M-S-1)(R-S-2)A + B\{(M-S)$$
$$(M-S+1)(M-R-1) + BC\} - (M-S)^3(M-S-1)]/D, \qquad (2)$$

where $A = \Sigma m^3 - 3\Sigma m^2 + 3M - S$,

$B = \Sigma m^2 - 2M + S$,

$C = 3M + 3R - 6S - 3 - 2(M-S)(R-S)$,

$D = 2(M-S)^2(M-S-1)^2(M-S-2)(M-S-3)$.

These values then yield the normal statistic $z = [T - E(T)]/[V(T)]^{\frac{1}{2}}$ for which the right-hand tail area probability may be derived to give the associated significance level. The mean and variance as calculated above are appropriate for data where each house contains at least one case of disease (i.e. where single ascertainment applies), and where all such households are included.

A second analysis which was carried out was to investigate the relationship of household size to the secondary attack rate within households. Primary and secondary cases were easily distinguished since the serial interval (incubation period plus the duration of the pre-eruptive stage) is sufficiently large. There were very few instances of multiple introductions and the interval between onsets of primary cases was so small as to leave no reasonable doubt about the intra-household generation to which the cases belong. The data were divided according to the household environment, and secondary attack rates were calculated for each household size, again allowing for the index case of each house. When a sample of households of given size was too small, some pooling of households of different sizes was used to create attack rates with a little more stability; this was usually necessary only for rather large households. Houses with only one resident cannot be used in this analysis, and are therefore excluded; in these data there were three such houses, and 182 houses with two or more residents were used.

Table 1. *Household aggregation of variola minor, by environment and phase of the epidemic growth*

Environment and/or Phase	R	Σr^2	S	M	Results of test for clustering						
					Σm^2	Σm^3	T	E(T)	$\sqrt{V}$(T)	z	P
All data	459	1599	185	919	5463	37549	570	487·8	14·6	5·64	< 0·001
Phase I	14	40	6	34	228	1606	13	13·1	1·7	-0·07	N.S.
Phase II	391	1391	153	789	4811	33687	500	430·0	13·7	5·11	< 0·001
Phase III	34	118	14	56	274	1628	42	34·8	4·1	1·78	< 0·05
Phase IV	20	50	12	40	150	628	15	10·0	1·3	3·82	< 0·01
Urban	152	440	71	387	2445	17361	144	127·4	6·6	2·49	< 0·01
Semi-rural	200	666	82	389	2253	15395	233	209·9	9·9	2·34	< 0·01
Rural	107	493	32	143	765	4793	193	165·9	8·0	3·37	< 0·01
Urban, Phase I	12	36	5	27	179	1263	12	11·9	1·6	0·06	N.S.
Urban, Phase II	140	404	66	360	2266	16098	132	115·3	6·3	2·63	< 0·01
Urban, Phases III and IV	No data										
Semi-rural, Phase I	2	4	1	7	49	343	1	Insufficient data to test.			
Semi-rural, Phase II	179	619	68	341	2057	14453	220	207·2	9·9	1·30	N.S.
Semi-rural, Phase III	11	33	6	19	65	241	11	Insufficient data to test			
Semi-rural, Phase IV	8	10	7	22	82	358	1	Insufficient data to test			
Rural, Phase I	No data										
Rural, Phase II	72	368	19	88	488	3136	148	129·9	6·3	2·87	< 0·01
Rural, Phase III	23	85	8	37	209	1387	31	29·7	3·7	0·34	N.S.
Rural, Phase IV	12	40	5	18	68	270	14	Insufficient data to test			

Table 2. *Relation of the secondary attack rate (SAR) of variola minor to household size, by environment*

| | Environment | | | | | | | |
| | Urban | | Semi-rural | | Rural | | Whole county | |
Household size*	No. †	SAR (%)	No. †	SAR (%)	No. †	SAR (%)	No. †	SAR (%)
2	0	—	6	25	2 ⎫ 61		8	29
3	11	23	12	44	8 ⎭		31	43
4	17	23	13	41	9	59	39	38
5	8	50	11	48	5	60	24	51
6	11	25	12	17	2 ⎫		25	30
7	8	21	8	60	2 ⎪		18	46
8	9	22	2 ⎫		1 ⎬ 78		12	31
9	3 ⎫		3 ⎪		1 ⎪		7	27
10	0 ⎪ 22		1 ⎬ 31		1 ⎭		2 ⎫	
11	1 ⎬		1 ⎪				2 ⎬ 28	
12	1 ⎭		1 ⎭				2 ⎭	

* Number of members without previous variola.

† Number of households with the given size (number of susceptibles).

RESULTS

Test for household aggregation

Considering the data from all 185 houses simultaneously, the following quantities were enumerated: $R = 459$, $\Sigma r^2 = 1599$, $S = 185$, $M = 919$, $\Sigma m^2 = 5463$ and $\Sigma m^3 = 37549$. This yields, after substitution into (1) and (2), that $T = 570$, $E(T) = 487\cdot8$, $V(T) = (14\cdot58)^2$, and hence $z = 5\cdot64$ indicating very significant clustering ($P < 0\cdot001$). When the study households are divided according to the phase or stage of the study epidemic, a significant clustering is noted in Phases II, III and IV but not in Phase I (Table 1). This probably results from the small number of households affected during Phase I. A similar finding is made when the households are divided by environment (Table 1), there being significant clustering in all three broad environments, urban, semi-rural, rural. When the study households are divided by both environment and phase of the epidemic growth (Table 1), small numbers appear and no definite trend is evident.

Secondary attack rates

Table 2 shows the relation of the secondary attack rate to the household size. Little trend is evident from the data for all houses combined. When the study households are divided according to the environment, urban and semi-rural households do not show any evident trend of the relation of the secondary attack rate to household size. Only rural households show a marginal trend (not significant), with larger houses showing the suggestion of greater susceptibility. This is possibly because larger households tend to contain more children.

DISCUSSION

Household size does not appear to be important. It seems that clustering has occurred generally, the significance of which is determined largely by the amount of data in each subgroup considered.

In general, the test is used within a given subset of data and it cannot be used directly to see, for example, if the strength of clustering is greater in urban than in rural households. An indirect comparison is possible through the P values but these are affected by the sample size as well as by the degree of aggregation present. In general, the P values here showed a consistent value when either the phase of the epidemic growth or the environment were examined, and all the results significant or not-significant (the latter usually occurring in small subgroups) were in the direction consistent with transmission between household members. In other words the data were consistent with household aggregation of variola minor rather than towards an excess spreading-out of the cases. In this regard, this dispersion of cases can also give significant results if extreme enough.

Any clustering in households which was evidenced can be interpreted as an indication of differential risks between households caused by some factors such as vaccination level (frequency of vaccinated members), school attendance status composition (relative frequency of adults and pre-school and school children) of the household and perhaps other factors. The clustering might also indicate greater 'contact' or potential for transmission between members of the same household than between members of different households.

REFERENCES

ANGULO, J. J. (1976). Variola minor in Braganca Paulista County, 1956: overall description of the epidemic and of its study. *International Journal of Epidemiology* 5, 359–65.

ANGULO, J. J. RODRIGUES-DA-SILVA, G. & RABELLO, S. I. (1967). Spread of variola minor in households. *American Journal of Epidemiology* 86, 479–87.

ANGULO, J. J., RODRIGUES-DA-SILVA, G. & RABELLO, S. I. (1968). Sociological factors in the spread of variola minor in a semi-rural school district. *Journal of Hygiene* 66, 7–18.

ARNT, N. & MORRIS, L. (1972). Smallpox outbreak in two Brazilian villages: epidemiological characteristics. *American Journal of Epidemiology* 95, 363–70.

DIXON, C. W. (1948). Smallpox in Tripolitania, 1946: an epidemiological and clinical study of 500 cases, including trials of penicillin treatment. *Journal of Hygiene* 46, 351–77.

HEINER, G. G., FATIMA, N. & McCRUMB, F. R., Jr. (1971). A study of intrafamilial transmission of smallpox. *American Journal of Epidemiology* 94, 316–26.

HENDERSON, R. H. & YEPKE, M. (1969). Smallpox transmission in southern Dahomey: a study of a village outbreak. *American Journal of Epidemiology* 90, 423–8.

KLAUBER, M. R. & ANGULO, J. J. (1974). Variola minor in Braganca Paulista County, 1956: space-time interactions among variola-minor cases in two elementary schools. *American Journal of Epidemiology* 99, 65–74.

PEARSON, K. (1912). On the appearance of multiple cases of disease in the same house. *Biometrika* 8, 404–12.

TROUP, J. M. & MAYNARD, G. D. (1912). Note on the extent to which the distribution of cases of disease in houses is determined by the laws of chance. *Biometrika* 8, 396–403.

WALTER, S. D. (1974). On the detection of household aggregation of disease. *Biometrics* 30, 525–38.