The distribution of episodes of illness—a research tool in general practice?

S. J. KILPATRICK, PhD.*
Research Unit, Royal College of General Practitioners, Birmingham

SUMMARY. The frequency of episodes of illness which 315,000 people brought to their doctors' attention in one year is shown to follow a geometric distribution. Some of the potential applications of this distribution in general-practice research are given.

Introduction

For the purposes of the second national morbidity survey, any face-to-face contact between doctor and patient is counted as a consultation. An episode of illness, on the other hand, is defined as a period of sickness for which there were one or more consultations. Episodes therefore reflect the number of ' problems ' seen by the general practitioner.

The frequency distribution of all episodes of illness recorded in the national morbidity survey for the year 1971-72 is given in table 1. This shows the number and percentage of people in the survey with 0, 1, 2 . . . episodes of illness. About 40 per cent of the 300,000 individuals surveyed did not see their doctor during the year, while about 1,500 (0·5 per cent) were treated for ten or more episodes of illness by their family physician.

TABLE 1
DISTRIBUTION OF EPISODES OF ILLNESS
(Excluding prophylactic procedures and routine antenatal and postnatal care)

Number of episodes	*Frequency*	*Frequency (%)*	*Percent predicted*
0	127,569	*40·4*	*40·0*
1	71,261	*22·6*	*24·0*
2	46,656	*14·8*	*14·4*
3	28,716	*9·1*	*8·7*
4	16,997	*5·4*	*5·2*
5	10,336	*3·3*	*3·1*
6	5,950	*1·9*	*1·9*
7 - 9	6,427	*2·0*	*2·2*
10 - 14	1,467	*0·5*	*0·6*
15+	110	*0·0*	*0·0*

(a)	Total number of episodes	=	473,426
(b)	Total number consulting	=	187,920
(c)	Total number of people at risk	=	315,489
(a)/(b)	Episode rate per patient consulting	=	2·5193
(a)/(c)	Episode rate	=	1·5006

(Source: Second National Morbidity Survey, table 9, males and females; Office of Population Studies and Censuses and the Royal College of General Practitioners, 1974)

Underlying distribution

The percentage frequencies given in table 1 yield a straight line relationship when plotted against the number of episodes on semi-logarithmic graph paper (figure 1). This means that the ratio of successive episode frequencies is constant. The only discrete statistical distribution which has this property is the geometric distribution (Ord, 1967).

*Honorary Research Fellow, Royal College of General Practitioners, current address: The Medical College of Virginia, Virginia Commonwealth University, U.S.A.

Journal of the Royal College of General Practitioners, 1975, **25**, 686–690

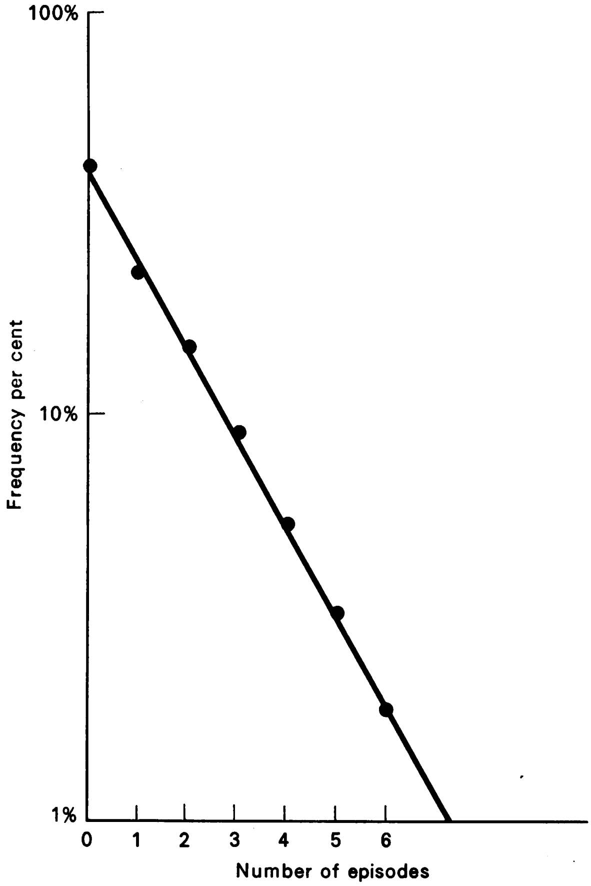

Figure 1.

If f(e) represents the frequency of individuals with e episodes, the geometric distribution specifies that

$$f(e) = (1-q)q^e \qquad (e = 0, 1, 2 \ldots)$$

where q is the constant ratio between successive frequencies,

$$q = f(e+1)/f(e)$$

Here (1–q) is the frequency of individuals with zero episodes. From table 1 this frequency is 40·4 per cent. A more efficient method of estimating q is to set

$$q = m/(1+m)$$

where m is the episode rate defined as the total number of episodes divided by the total number of persons at risk. From table 1, the episode rate (m) is 1·50, giving an estimate of q equal to 60·0 per cent.

In symbols, given

$$m = 1{\cdot}50$$
$$q = 1{\cdot}5/2{\cdot}5$$
$$= 0{\cdot}600$$
$$\text{and } 1-q = 0{\cdot}400$$

Although slightly lower than the observed zero frequency of 40·4 per cent, this estimate uses all the information available. We observe a zero frequency of 40·4 per cent but expect a frequency of 40·0 per cent if the geometric distribution holds.

The observed frequency of individuals with one episode is 22·6 per cent. The geometric distribution predicts the frequency of single episodes to be

$$\begin{aligned} f(1) &= (1\text{–}q)q \\ &= 0{\cdot}400 \times 0{\cdot}600 \\ &= 24{\cdot}0 \text{ per cent} \end{aligned}$$

The predicted frequency for two episodes is

$$\begin{aligned} f(2) &= (1\text{–}q)q^2 \\ &= 0{\cdot}400 \times 0{\cdot}600 \times 0{\cdot}600 \\ &= 14{\cdot}4 \text{ per cent} \end{aligned}$$

as compared with an observed frequency of 14·8 per cent. These and the remaining frequencies predicted from a geometric distribution with $q = 0{\cdot}600$ are given in table 1. Excellent agreement is obtained between the frequencies observed and those predicted.

In figure 1, a straight line representing the geometric distribution has been drawn through the predicted frequencies. It is interesting how closely the observed frequencies follow this line.

Discussion

Froggatt *et al.* (1969) found that the negative binomial best describes the distribution of consultations in a Belfast general practice. The distribution which episodes follow (the geometric distribution) is a simpler form of the negative binomial.

Some applications in general practice of the finding that episodes follow a geometric distribution can now be considered.

Epidemiology

The most important of these is the inference that episodes of illness are the result of a random process. This follows from the finding that the chance of an additional episode is independent of the previous number of episodes.

That episodes of illness are generated by a random mechanism is rather surprising, considering the known heterogeneity of reported morbidity and the aetiological chains and temporal clusterings among many morbid conditions. Clearly, further investigation is required into how a collection of dissimilar diseases can give such a regular progression as the geometric distribution when combined.

Characterisation

By custom, the morbidity in a practice in a given period is measured by the episode rate per patient consulting, i.e.

$$m' = \frac{\text{Total episodes}}{\text{Total patients consulting}}$$

or by the episode rate, i.e.

$$m = \frac{\text{Total episodes}}{\text{Total persons at risk}}$$

The choice of m' and m as summary statistics is, however, arbitrary since they relate to no underlying distribution in the way that (say) the mean relates to the normal distribution (Kilpatrick, 1973). The constant ratio, q, between successive episode frequencies, may be used in place of m' and m since q fully specifies the distribution of episodes. However, since

$$m' = \frac{1}{1\text{–}q}$$

and

$$\begin{aligned} m &= m'q \\ &= \frac{q}{1\text{–}q} \end{aligned}$$

it is now easy to pass from one summary statistic to another.

For sufficiently large and homogenous populations at risk we may therefore characterise their morbidity as seen in general practice by either of the above episode rates or by q, the proportion who bring at least one episode of illness to their doctor's attention in the period. Groups defined by age, sex, income, occupation, religion, race, place of residence, or by the doctor or practice they register with, could be typified from national data.

Prediction

Given the values of q for groups defined by each of the factors above, it should be possible to predict the q or episode rates to be expected from a group by combinations of the factors—perhaps even to the extent of predicting the expected distribution of episodes from each practice or from specified cohorts. Indeed according to the theory of the negative binomial (Froggatt *et al.*, 1969) there should be perfect correlation between the distribution of episodes in successive periods.

Departures from the geometric distribution

If it is assumed that a large homogenous group will generate episode frequencies which lie on a straight line when plotted on semi-logarithmic paper, then consistent deviations from this straight line will reveal the influence of other factors. For example, a potential source of bias in calculating episode rates is due to recent arrivals. It is thought that when a family moves to another neighbourhood, there is little motivation to register with a local general practitioner until a member of the family needs to consult him. The effect of such delayed registration would be to decrease the number at risk, thus inflating the value of the episode rate m so that

$$m > m'q$$

In fact, from table 1

$$m = 1{\cdot}5006$$

while

$$m'q = 1{\cdot}5118$$

showing, if anything, the opposite effect. We cannot then attribute this difference to delayed registration, which in any case should be largely balanced by persons who have left the district, but still registered with their last general practitioner.

The difference may be due to multiple registration although care has been taken to remove multiple registrations from the practice list of this survey where they have been detected.

Similarly, the difference between observed rates or frequencies and those expected, by compounding the values of q for the different factors which characterise the group, could be used as a test of independence of these factors. For example, it is unlikely that age and sex can be treated as separate factors acting independently.

Truncation

One reason to use the episode rate per patient consulting is that the number not consulting is estimated inaccurately (as in Canada) or is unknown as in the United States.

However, either the truncated distribution of episodes, excluding zero episodes or the episode rate per patient consulting, m′, will be known. From the above relationships it may be shown that

$$m = m'q = m'-1$$

i.e. the episode rate is the product of the episode rate per patient consulting and q, the constant proportion between frequencies or, equivalently, is equal to one less than the episode rate per patient consulting.

If n′ is the total number of patients consulting and n is the total number registered or at risk then

$$n = \frac{n'}{q} = \frac{n'm'}{n'-1}$$

i.e. the total practice size is the number consulting divided by q or, equivalently, the total number of episodes (n′m′) divided by one less than the episode rate per patient consulting. Finally, we may estimate n″, the number who do not consult as

$$n'' = \frac{n'\ (1-q)}{q} = \frac{n'}{m'-1}$$

i.e. the number not consulting is the number consulting divided by one less than the episode rate per patient consulting.

It is intended to explore these and other implications of the geometric distribution of episodes and to demonstrate their application to real situations in subsequent publications.

Acknowledgements

I am indebted to Dr D. L. Crombie for his suggestions and encouragement and to Mrs J. Mant for typing this manuscript. This work was partly supported by N.I.H. Grant HS 01899-01.

REFERENCES

Froggatt, P., Dudgeon, M. Y. & Merrett, J. D. (1969). *British Journal of Preventive and Social Medicine*, **23**, 1–11.

Jacob, A. & Pearson, J. (1967). *Journal of the College of General Practitioners*, **13**, 303–312.

Kilpatrick, S. J. (1973). *Statistical Principles in Health Care Information*. P. 80, Baltimore: University Park Press.

Office of Population Studies and Censuses and the Royal College of General Practitioners (1974). *Morbidity Statistics from General Practice*. London: H.M.S.O.

Ord, J. K. (1967). *Journal of the Royal Statistical Society* A, **130**, 232–238.

HEALTH CARE IN SCOTLAND

The report for 1974 on Health Services in Scotland states that cigarette smoking is the largest single avoidable danger to health in Scotland today. This country spends more per household and a larger percentage of total weekly expenditure on tobacco than any other area of Britain. The penalty it pays for this is shown by the health statistics. In 1973, nine per cent of deaths among males were caused by cancer of the bronchus and lungs.

The report describes Scotland as the most toothless nation in the world. In a recent survey 44 per cent of Scots aged 16 and over had lost all their natural teeth and only about two per cent of the population could be classified as dentally fit.

Recent analysis had shown that 20 per cent of admissions to mental hospitals were for patients aged 65 years and over and 48 per cent of patients resident in these hospitals were in this age group

The average cost for outpatient treatment had risen to £2·01 in 1974.