Frequency distribution of hospital-referred parasuicidal episodes in Edinburgh

JOHN C. DUFFY

From the Department of Statistics, University of Edinburgh, and MRC Unit for Epidemiological Studies in Psychiatry, Royal Edinburgh Hospital, Edinburgh

SUMMARY This paper presents frequency distributions of parasuicidal episodes (suicide attempts) in Edinburgh for four subgroups of people for each of the years 1970, 1973, and 1974. Probability distributions are fitted to these data and it is shown that the log-series distribution best fits the data for two subgroups. The stability of the parameters of these distributions is examined and is demonstrated for two of the subgroups.

Since 1968 data on parasuicide referrals to the Regional Poisoning Treatment Centre (RPTC) of the Royal Infirmary at Edinburgh have been recorded routinely. This information has been used extensively in studies of parasuicide in Edinburgh, most recently by Holding et al. (1977). The definition of parasuicide in this paper conforms to that of previous studies in that the phrase 'parasuicidal episode' is taken to mean a non-fatal deliberately initiated act of self-poisoning or self-injury.

Some probabilistic aspects of the phenomenon are examined to evaluate observed differences in rates both within Edinburgh (for example, in different years) and between different populations. The work is relevant to the measurement of repetition and to the selection and testing of possible stochastic models of parasuicide. This paper is not concerned with psychosocial explanations of parasuicide in the individual case, but rather with the pattern of parasuicide viewed as a mass phenomenon.

These models may point to aspects in the behaviour of individuals against which possible theories may be examined so that the validity and effectiveness of standard statistical techniques applied to measurements of the phenomenon can be assessed.

This work attempts to find a single parameter probability distribution which best fits the observed frequency distributions. Attention is restricted to single-parameter distributions because of the small number of frequency classes in the data. Three distributions are considered: the Poisson, geometric, and logarithmic series (log-series), which are the most frequently encountered single parameter distributions in a variety of applications. Complete details of these and other discrete distributions are given in Johnson and Kotz (1969) but for simplicity of exposition their stationary forms are described here. It may be noted however that each distribution may be generalised to admit certain forms of time dependence in the appropriate parameter.

Distributions

POISSON

This distribution would be expected to fit observed data if all individuals were equally likely to sustain an episode in a given period of time, and if for each individual the probability of a future episode was independent of the number and timing of episodes in the past. The Poisson distribution has been applied to many random phenomena—for instance, to naval aviation accidents, Supreme Court vacancies in the USA, and parasuicide in small areas of Oxford (Skrimshire, 1976). It has been found to apply to many classes of data arising from events which occur randomly in time and is particularly important in the mathematical study of queues. The distribution is often uncritically assumed in order to facilitate statistical analysis.

The mathematical definition of the distribution is given by

\[ P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \]

\[ x = 0, 1, \ldots ; \lambda > 0 \]

—that is (in the present case), the probability that an individual sustains \( x \) episodes in a unit period of time is given by the expression on the right hand side, which involves only one unknown parameter, \( \lambda \).
GEOMETRIC
This distribution might be expected to apply to parasuicide if each individual in the population had a unique risk of parasuicide in a given period of time, different for different individuals (but constant and independent of the number and timing of previous episodes in a particular individual), and distributed according to the exponential law with parameter $1 - p$. It is a special case of a two-parameter distribution, the negative binomial distribution which is mentioned later but not considered in detail. The geometric distribution has been applied to problems in queuing theory and meteorology.

The defining relation may be written

$$P(X = x) = p(1 - p)^x$$

$$x = 0, 1, \ldots ; 0 < p < 1,$$

$p$ being the unknown parameter.

LOG-SERIES
This distribution may be generated in two simple ways. In common with the preceding two it may be expected to apply when in the individual case the risk of a further episode is independent of the number and timing of previous episodes, but where the individual risks for different individuals follow a particular distribution (the truncated gamma distribution) in the population; alternatively it arises from a process in which the initial individual risks are constant but this risk increases with number of episodes. It is not possible to discriminate between these geneses by univariate goodness of fit tests, rather the joint frequency distribution of the number of episodes sustained in two consecutive periods of time is necessary to examine the adequacy of these models. The distribution has found applications in biology, meteorology, and economics.

The mathematical definition of the distribution is

$$P(X = x) = \alpha \theta^x/x$$

$$x = 1, 2, \ldots ; 0 < \theta < 1;$$

$$\alpha = -\log(1 - \theta))^{-1}$$

which involves $\theta$ as the unknown parameter.

Method

The data used here are essentially those reported by Holding et al. (1977). They consist of the records of all parasuicide referrals (episodes) to the RPTC between 1968 and 1974, of those individuals aged 15 years or more and resident in Edinburgh.

It was necessary to link records pertaining to the same individual in order to construct the frequency distributions, and this task was complicated by the absence from the records for reasons of confidentiality of any information capable of uniquely identifying an individual by machine processing. Such linkage had therefore to be carried out manually, and for reasons of availability of records had to be within-year links for the years 1970, 1973, and 1974. Errors in the data file discovered during the manual linkage were corrected and figures quoted in this paper are therefore in places slightly different from those in other studies of the same data.

Previous studies have reported differences in repetition between men and women, and between those with and without previous episodes (Kessel and McCulloch, 1966; Buglass and Horton, 1974). To take account of these differences separate analyses were performed for the four subgroups of persons: men/women, first-ever/non-first-ever (or repeaters). The last term describes persons who have had one or more parasuicidal episodes earlier than the calendar year being considered.

Frequency distributions of the number of individuals sustaining $x$ episodes ($x = 1, 2, \ldots$) within each of the three years of study were tabulated for each of the four subgroups. The overall admissions : patients ratios for these years were calculated separately by sex.

An attempt was made to determine the stability of the mean number of episodes for each subgroup during the years 1971 and 1972. The numbers of men and women first-eters and repeaters for these years were ascertained and multiplied by the mean number of episodes (for the appropriate subgroups) derived from the data for 1970, 1973, and 1974. The total 'expected' number of admissions was then obtained by addition and compared with the actual number. Close agreement would be evidence in favour of a stable pattern and would lend confidence to further analyses.

The three probability distributions described earlier were fitted to the data for 1970, 1973, and 1974 and some comparisons and goodness of fit tests were performed. The Poisson and geometric distributions were fitted in their zero-truncated forms given by

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!(1 - e^{-\lambda})}$$

$$x = 1, 2, \ldots ; \lambda > 0$$

and

$$P(X = x) = p(1 - p)^{x-1}$$

$$x = 1, 2, \ldots ; 0 < p < 1$$

respectively.

The zero-truncated distributions were used because of the difficulty of obtaining an appropriate
Frequency distribution of hospital-referred parasuicidal episodes in Edinburgh

total population figure which would allow completion of the zero class in the frequency distributions. For the first-ever group it is arguable that the total population of Edinburgh would be the appropriate figure because the proportion in this population with previous episodes (and therefore not at risk of entering the first-ever frequency distribution) is negligible. In the light of the data it is clear that to use such a figure will result in an excess of zero observations in comparison with the fitted distributions. This problem may be overcome by assuming an unknown proportion of 'immunes' in the population and then estimating this proportion. The latter course leads to the same parameter estimates as are obtained by fitting the zero-truncated distributions.

Results

The Stability of the Frequency Distributions

Table 1 shows the frequency distributions constructed from the data for 1970, 1973, and 1974.

It will be seen that the number of individuals sustaining four or more episodes is small in the first-ever groups, but that this is not so in the repeater groups. The mean number of episodes in the repeater groups is consistently higher than that for the corresponding first-ever groups. In the first-ever groups the mean is stable for women, slightly less so for men, and there appears to be little difference between the sexes on this variable. In the repeater groups the mean for women is again stable but that for men is not, being high in 1973.

The admissions: patients ratio has been used in the past as an approximate measure of repetition within a calendar year. Table 2 shows this ratio for the target years. It can be seen that the ratio does reflect to some extent the amount of repetition but that it is insensitive to the differing proportions of first-ers in the total groups. The usefulness of this ratio will be discussed later.

Table 3 shows the results of estimating the 1971 and 1972 totals on the basis of the means for the target years. They approximate the actual number of admissions quite closely for both sexes. A slight
underestimation of the total number of admissions was to be expected, since, as mentioned earlier, errors were discovered during construction of the frequency distributions and of these the two most common were the recording of a repeater as a first-ever and the duplication of an admission. It seems likely, therefore, that the means of the frequency distributions in 1971 and 1972 are not greatly different from those for the target years.

FITTING THE DISTRIBUTIONS

Table 1 gives the expected frequencies calculated on the basis of the Poisson, geometric, and log-series distributions. The maximum likelihood estimates of the parameters and values of the $\chi^2$ goodness of fit statistics are also given. None of these $\chi^2$ values attains statistical significance and therefore there are no statistical grounds for discriminating between the three distributions in the first-ever groups.

It is not possible to perform a similar test on the repeater groups because of the possibility of the same individuals appearing in the tables in different years, thus vitiating the assumption of independence essential to the statistical basis of the test. However it can be seen that the log series distribution best fits the frequency distributions for each repeater group. In each year for either sex the value of the $\chi^2$ statistic is least when this distribution is applied. The poor fit of the truncated Poisson distribution is particularly striking.

Table 4 gives the estimated number of individuals with no episodes in each year calculated by completing the truncated Poisson and geometric distributions fitted to the first-ever frequency distributions. These figures would notionally represent the number of individuals in the population with no previous episodes and who sustained no episodes in the year in question but were 'at risk'.

Table 2 Admissions: patients ratios

<table>
<thead>
<tr>
<th>Sex</th>
<th>1970</th>
<th>1973</th>
<th>1974</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1:14</td>
<td>1:31</td>
<td>1:18</td>
</tr>
<tr>
<td>Women</td>
<td>1:13</td>
<td>1:20</td>
<td>1:20</td>
</tr>
</tbody>
</table>

Table 3 'Expected' and actual episodes in 1971 and 1972

<table>
<thead>
<tr>
<th>Sex</th>
<th>1971</th>
<th>1972</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Persons</td>
<td>Expected</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-ers</td>
<td>214</td>
<td>229</td>
</tr>
<tr>
<td>Repeaters</td>
<td>134</td>
<td>201</td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>430</td>
</tr>
<tr>
<td>Actual no. of episodes</td>
<td>437</td>
<td>443</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-ers</td>
<td>344</td>
<td>368</td>
</tr>
<tr>
<td>Repeaters</td>
<td>236</td>
<td>295</td>
</tr>
<tr>
<td>Total</td>
<td>580</td>
<td>663</td>
</tr>
<tr>
<td>Actual no. of episodes</td>
<td>674</td>
<td>762</td>
</tr>
</tbody>
</table>

They are very much smaller than the total populations since in 1971 there were over 350000 people in Edinburgh over 15 years of age.

Discussion

It is necessary to emphasise that the data derive solely from parasuicidal episodes which result in admission to the RPTC. It has been shown (Kennedy and Kreitman, 1973) that there are no systematic differences on a number of important variables between those episodes which are treated solely by general practitioners and those which result in admission to the RPTC, but that the proportion of episodes resulting in admission varies inversely with the number of earlier episodes in the history of the individuals concerned. The pattern of repetition manifest in admissions to the RPTC will therefore tend to underestimate the situation. Consequently, the results presented and their implications cannot be uncritically extended to all parasuicidal episodes in Edinburgh.

It has been shown that of the distributions considered the log-series distribution best fits the frequency distribution of the number of episodes in the repeater groups. The fitting of probability distributions has on occasion led to erroneous inference in other fields. As Arbous and Kerrich (1951) pointed out, the model of 'proneness' or unequal liability of different individuals to sustain certain events leads under certain assumptions to univariate distributions which may also arise from models postulating a risk increase resulting from each event sustained (contagion). The simplest way to decide which of these two models provides the best description of the phenomenon is to construct a bivariate frequency distribution of the number of episodes in two consecutive periods of time, for in the bivariate case the models lead to distributions with different parameters. The log series distribution may arise in both these ways as a limiting form of another distribution, the negative binomial distribution*. The proneness interpretation of the log

*The truncated negative binomial distribution gives a good fit to the frequency distribution in Table 1 when first-ers and repeaters are combined. The distribution also fits well to similar data from Oxford (1972-73) obtained from Bancroft (1974) personal communication, and from Malmo (1968, 1969) given in Linne (1972). In every case the exponent estimate is negative and the resulting complete distribution is improper, making estimation of the zero class impossible.
series distribution has been used in studies of consumer purchasing (Chatfield et al., 1966) in which the distribution was found to give a good fit to the distribution of the number of consumers making various numbers of purchases during a given period of time. The proneness model was tested on the basis of prediction of further purchasing in a consecutive period of time and found to give an adequate fit. This application is analogous to that considered here and suggests a further stage of analysis based on records linked between years, in other words a bivariate approach.

The contagion model, which postulates that the likelihood of repetition increases with the number of attempts made by an individual, leads to exactly the same univariate distribution as does the ‘proneness’ model and for this reason the current work is unable to discriminate between these. In fact it is not necessary that either model should fit the phenomenon; the empirical goodness of fit of the log series distribution to data within periods of one year does not guarantee the adequacy of the distribution to describe the phenomenon during time intervals of different length.

The first-ever groups show no evidence of either heterogeneity or contagion, an equal constant risk model (that is, the Poisson distribution) giving an adequate fit. It appears therefore that at some stage (possibly after the first-ever episode) there is a change in an individual’s risk and that this change is not the same for different individuals. This might be tested by a follow-up study to compare first-evers with matched controls who have no record of parasuicide. There is no evidence of a consistent increase in risk after each subsequent episode in an individual’s record, although a more stringent test of this would be afforded by records linked between years.

There are other important consequences of this work for the interpretation and analysis of parasuicide data. There are several statistical techniques which attempt to predict an outcome variable on the basis of measurements of associated variables, many nowadays capable of dealing with large numbers of predicting variables and with outcome variables of many types. These methods usually involve explicit distributional assumptions which may now be examined. Skrimshire applied such a method to the prediction of the number of parasuicidal episodes in small areas of Oxfordshire on the basis of social and demographic characteristics of the areas. The method makes the assumption that the Poisson distribution applies to the number of episodes for each individual. It has been shown here that this is not so, and further that there will be a discrepancy between the observed and expected number of zeros which will lead to a poor fit, with underestimation of multiple repetition. This was noticed by Skrimshire (1976), whose explanation has now been confirmed. Table 4 shows that for the Poisson or geometric distributions to apply the estimated population at risk is very much smaller than the total population, hence to apply these distributions to parasuicide one must postulate the existence of a group who are not at risk (immune) and who constitute a very large proportion of the population. The estimation of this proportion should then form part of the statistical analysis. The concept of ‘immunity’ in this case is introduced purely as a mathematical convenience; while it is not impossible that there exist persons who are not at all at risk of parasuicide their existence is in no way established by the current work.

Much work in the field of parasuicide consists of the presentation and interpretation of observed rates. While these are important for summary description of the phenomenon their relationship with the processes described earlier requires some clarification. In the work of Holding et al. (1977) a first-ever rate, a patient rate, and an admission rate were all employed. These rates were calculated on an annual basis thus facilitating comparisons between years.

The first-ever rate is calculated by dividing the total number of first-ever episodes in a particular year by the total population of the geographical region in question (in this case the city of Edinburgh) and multiplying by an appropriate constant (for example, 100 000 to give a rate per 100 000). First-ever rates for different years may be compared in a simple fashion since such rates are examples of the binomial rates described in elementary textbooks on medical statistics. Rates for different years may be regarded as statistically independent if the slight changes in the denominator owing to first-ever episodes in intervening years are ignored.

The patient rate is calculated as above, with the numerator defined as the number of individuals sustaining one or more parasuicidal episodes in the appropriate year. It therefore contains the first-ever rate, the remainder being made up of those individuals sustaining one or more episodes in the year who have a history of parasuicide in previous years.

Patient rates for different years may involve the same individuals and since it has been shown here that the population is not homogeneous with respect to the probability of repetition they are not statistically independent. Comparisons of this rate between different years are therefore limited in their interpretative value. If an attempt
is made to adjust for differences in the first-ever rate by, say, considering only those individuals who have a history of parasuicide before the year in question (the 'remainder' mentioned above) it is clear that only those individuals in the population at large with a previous history of parasuicide are 'at risk' of being counted in this total. The number of such individuals will usually be unknown and hence a rate cannot be calculated. If the number of such individuals sustaining episodes declines in comparison with previous years it follows that the rate of repetition in the population has fallen. It is not however conversely true that an increase in this number implies an increase in the rate of repetition, since the number of individuals at risk of repetition increases each year as a result of the first-ever episodes in the previous year.

The admission rate is defined similarly to the above two with the numerator given by the total number of episodes in the year in question. Subsumed in it are the two rates previously considered and the total number of episodes associated with each. Some of the problems of interpreting this rate can be illustrated by considering the effect of a few simple assumptions on the change in the admission rate over time. In the case of a constant first-ever rate a constant number of episodes will be generated each year by individuals who have had no previous parasuicidal episode. Each year the total number of individuals at risk of repetition will increase by the addition of those who sustained their first-ever episode during the previous year. Assuming that a constant proportion of these individuals sustain an episode each year, and that the number of episodes for an individual is also constant, then it is clear that an increasing admission rate will be observed. In other words, in a situation where the population parameters remain constant, one of the indices used to describe the phenomenon—that is, the admission rate—is bound to increase. This example ignores population changes, death rates, etc., but it clearly indicates a need for caution in interpreting changes in the admission rate in terms of processes in the population.

The measurement of repetition is an important aspect of such work in this field. In the past the extent of repetition in various groups of patients has been measured in many different ways. The admissions : patients ratio quoted by Bancroft et al. (1975) and Holding et al. (1977), is perhaps the easiest quantity to calculate routinely, although its utility in an explanatory model is limited. A similar stricture applies to the percentage of patients repeating within a given follow-up period (Kessel and McCulloch, 1966; Buglass and Horton, 1974) and the percentage of patients repeating within the calendar year (Kessel and McCulloch, 1966). The chief difficulties of such quantities are that they are insensitive to multiple repetition and differential repetition rates within the cohort studies: these are the necessary consequences of using one parameter estimate in a situation which may require many for its description.

One of the first essentials for monitoring of repetition is a reliable estimate of the proportion of individuals in the population who have at least one episode of parasuicide in their history. A further necessity is a fully linked records system. In the absence of the first desideratum it is possible to examine the pattern of repetition within a given period of time for a group who have sustained at least one episode, and to compare this pattern and its parameters across different years. Such comparisons permit more detailed interpretations of observed differences in summary rates, and may suggest models of the phenomenon. When suitable models have been constructed and tested, more exact statistical analyses will be facilitated and a firmer empirical basis provided for explanatory theories of the processes underlying parasuicide.

I thank my colleagues in the MRC Unit for their advice and encouragement, particularly Norman Kreitman, Trevor Holding, and Dorothy Buglass. I am grateful to the referee for his constructive comments and suggestions.

Reprints from J. C. Duffy, MRC Unit for Epidemiological Studies in Psychiatry, Royal Edinburgh Hospital, Morningside Park, Edinburgh.

References


