A SIMPLE METHOD FOR THE DETERMINATION OF SAMPLE SIZE IN SURVEYS FOR INTESTINAL PARASITES

NEIL A. CROLL and TERRENCE D. SOLE

Institute of Parasitology, Macdonald College, McGill University, P.Q. Canada H9X ICO.

INTRODUCTION

Selection of a suitable sample size is critical for surveys of both human and animal parasites. The sample size must be large enough to account for virtually all the parasite species present in their relative proportions but small enough for practical considerations.

Cochran (1963) and Southwood (1966) have suggested that the sample size should be determined after a pilot survey and after analyzing the variance of the results obtained. While this is a mathematically vigorous approach, it is seldom possible. Since parasite populations are often overdispersed in their hosts (Crofton, 1971), calculations can become very complex.

THE SAMPLE SIZE

A method is proposed here in which the results are continually monitored to determine the point of redundancy. This is based on the ecological tenet that a community of any group of organisms has an unequal species distribution. Thus in a human population there is a small variety of common parasites which are found in many hosts and several parasites in very few hosts (Croll and Ghadirian, 1981). Furthermore, there will be an inability to find the increasingly rarer species. The rarest species may be missed even by examining all hosts, so a point of compromise is reached when the survey is judged to be representative.

Fig. 1 shows a theoretical relationship between the number of species found and the number of samples. Common species are soon discovered, while the sample size is small. As the curve approaches the asymptote, only the rarer species are added to the total. Preston (1948) has suggested that the distribution of species in a community can be described by a log-normal curve to various bases depending on the circumstances; therefore, the sample size would have to be increased exponentially to find new species, an increase which quickly results in enormous effort.

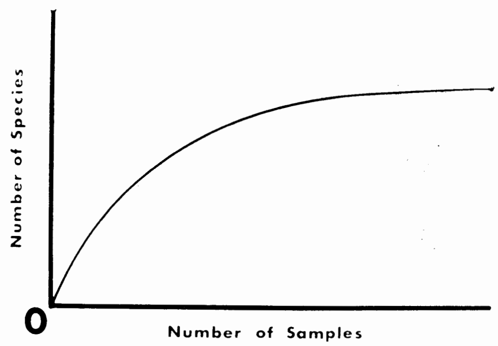

Fig. 1—The theoretical cumulative species curve. As the number of samples increases the total number of species increases. However, the rate of finding new species decreases and rare species necessitate huge sampling campaigns.

It should be possible to determine a point when most of the parasite species in a host population have been discovered and to estimate the number of rarer species. The monitoring of the "*Cumulative Species Curve*" for intestinal parasites will indicate an adequate sample size during the course of a survey in the following manner.

During a survey of random stool samples from a population, the number of species in the first stool sample is plotted and the number of additional species found in the second sample is added and so on. "New" species, found for the first time, in the third, fourth and subsequent samples are added to the total. As the number of samples increases the number of "new" species decreases and the curve approaches an asymptote and more host samples will not significantly increase the total number of species. A projection of the curve will give an accurate estimation of the diversity of the parasite population.

To test this method, data from human surveys by Cross *et al.*, (1976), Ghadirian (1978 pers. comm.) samples 190 persons in Iran and found 10 parasite species; Meerovitch and Gibbs (1969) examined 211 individuals on Easter Island and identified 12 species; and Cross *et al.*, (1976) found 10 different parasite species in a sample population of 164 persons from Indonesia. Could the information in these surveys have been obtained from fewer samples, or are these sample sizes actually too small and not representative of the actual diversity?

The raw data from each of these surveys were computerized. Random selections (without replacement) and without reference to age and/or sex were made using three different sets of uniformly distributed computer-generated random numbers.

Figs. 2a, b and c show the *Cumulative Species Curves* produced by the randomization of the data from three surveys as previously described. The data of Ghadirian (1978) (Fig. 2a) will be discussed further; the other surveys were treated identically.

In the first survey 19 persons demonstrated 8 species (Fig. 2a). The sample size could have been doubled and, if after 28 persons had been examined, a ninth species had not been found, then sampling could have stopped. In the 35th person a ninth species was identified. This value would then be doubled again and if a tenth species was not found by the 70th person, sampling should have ceased. A tenth species was found in the 49th person; therefore, the sample size would then become 98 persons. All ten species would have been found in the 98 persons. This provides a conceptual framework to be used as the sampling proceeds.

The predicted point of greater confidence for the data of Cross *et al.*, (1976) is 96 persons and for Meerovitch and Gibbs (1969) is 46 persons.

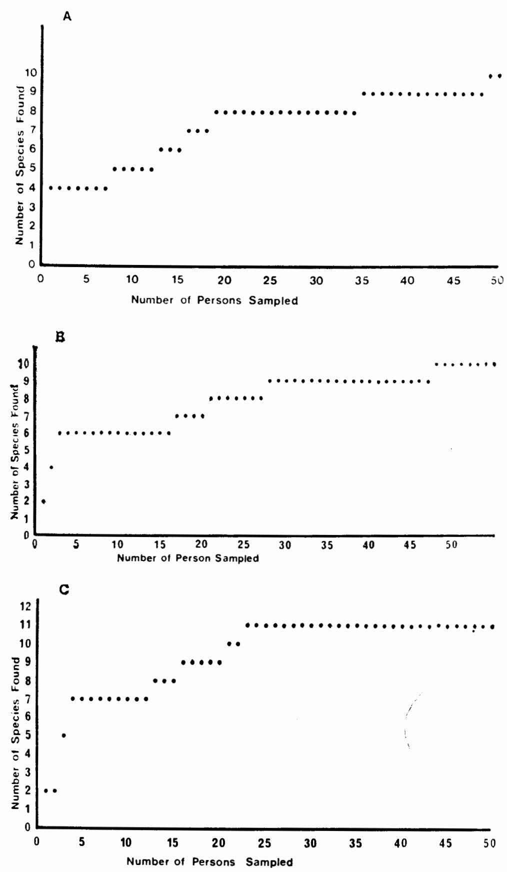

Fig. 2—Cumulative species curves produced by selecting the first 50 persons at random from the larger sample. *A*. Data from Ghadirian (1978), *B*. Cross *et al*, (1976) and *C*. Meerovitch and Gibbs (1969).

DISCUSSION

Traditionally, sample sizes of surveys for human intestinal parasite infections have been based on intuition, convenience, an arbitrary percentage of the total host population, or on a fixed number such as "100". There appears to be no particularly good reason to have a standard *a priori* since it might be wastefully large in some cases and uninformatively non-representative in others.

The method of determining the sample size in human parasitologic surveys is based on the application of the species abundance curves which have been extensively discussed in population ecology (Pielou, 1969, 1975; Preston, 1948). These studies suggest that in any community of organisms there are a few species which are very abundant and many more species which are less common. Intestinal parasites appear to conform to this pattern of diversity.

The use of the *Cumulative Species Curve* favours a sample size which is appropriate to the diversity of the community. In communities of low diversity (i.e. all hosts are infected with similar parasites) the sample size will be small, whereas host populations with high diversity (i.e. hosts have many different parasites) will require larger sample sizes. It also permits that those populations with low overall parasite prevalence may require larger samples than those with a higher prevalence.

As with any survey method very rare species may be missed. This point is well illustrated by the computerized random sampling of the data of Meerovitch and Gibbs (1969); *Cumulative Species Curve* indicated that sampling should cease after examination of 46 persons. However, even after 100 persons had been randomly selected the model failed to show *Isospora belli* which Meerovitch and Gibbs (1969) had found in 1 of 211 persons (0.47%).

If the goal of the survey is to identify the majority of the species present and to determine their relative proportions for epidemiological and health planning purposes, then the generally smaller sample sizes determined by the *Cumulative Species Curve* will suffice. These relative proportions of each parasite species in the hosts sampled should not be significanty different from the proportions of the total host population.

It is of particular interest that the total protozoal and helminthic parasites of the human gastrointestinal tract, which are identified in stool samples, conform to the basic ecological tenets of diversity which are known in other communities. Factors such as host age, sex, variation in parasite fecundity, concurrent infections and other complicating factors of the host do not appear to influence the overall pattern of diversity in the parasite community of the gastrointestinal tract.

SUMMARY

A method is proposed for the determination of a sample size in surveys for intestinal parasites, based on the ecological tenet that a community of organisms has an unequal species distribution.

ACKNOWLEDGEMENTS

The authors are particularly grateful to Dr. J.H. Cross, Dr. E. Ghadirian, and Dr. E. Meerovitch for permitting access to the raw data of their surveys. It is also a pleasure to acknowledge financial support from the National Science and Engineering Research Council of Canada, the Canadian International Development Agency and the World Health Organization.

REFERENCES

COCHRAN, W.G., (1977). *Sampling Techniques* 2nd Ed. John Wiley & Sons, Toronto.

CROFTON, H.D., (1971). *Parasitology*, *62* : 179.

CROLL, N.A. and GHADIRIAN, E., (1981). Wormy persons: contributions to the nature and patterns of overdispersion in persons with *Ascaris lumbricoides*, *Ancylostoma duodenale*, *Necator americanus*, and *Trichuris trichura*. *Trop. Geogr. Med.*, (in press).

CROSS, J.H., CLARKE, M.D., COLE, W.C., LIEN, J.C., PARTONO, F., DJAKARIA, JOESOEF A. and OEMIJATI, S., (1976). Parasitic infections in humans in West Kalimantan (Borneo), Indonesia. *Trop. Geogr. Med.*, *28* : 121.

MEEROVITCH, E. and GIBBS, H.C., (1969). Intestinal parasitic infections in the inhabitants of Easter Island. *Trans. Roy. Soc. Trop. Med. Hyg.*, *63* : 370.

PIELOU, E.C., (1969). *An Introduction to Mathematical Ecology*. John Wiley & Sons. New York.

PIELOU, E.C., (1975). *Ecological Diversity*. John Wiley & Sons, New York.

SOUTHWOOD, T.R.E., (1966). *Ecological Methods*. Chapman & Hall, London.