Anim. Behav., 1992, **44**, 377–379

SHORT COMMUNICATIONS

Measuring association between individuals or groups in behavioural studies

JOSHUA R. GINSBERG*† & TRUMAN P. YOUNG‡
**Department of Zoology, University of Oxford, South Parks Rd, Oxford OX1 3PS, U.K.*
‡Department of Botany, University of California, Davis, Davis, CA 95616, U.S.A.

(Received 2 April 1991; initial acceptance 15 May 1991;
final acceptance 15 October 1991; MS. number: AS-798)

Association indices have been used by ecologists, particularly in botanical studies, for many years. An understanding of how these indices were derived is critical to an understanding of why most of them are inappropriate for use in behavioural studies (see review in Cairns & Schwager 1987). In ecological studies, indices are used to measure association of two species in space: how often (in how many transect sites) do two species co-occur? In answering this question, an attempt is made to reconcile two different, but real values. For instance, a particular species of tree may be found only in the woodlands while a species of grass may be found in both woodlands and plains. Let us assume that an even number of sites are censused in each habitat. In all woodland plots, grass and tree co-occur, so the association of grass and tree from the perspective of the tree species is 100%. In all grassland plots, the grass is found without the tree, so the association of the grass with the tree is 50%.

This inequality can be expressed mathematically (definitions in Table I). In an index of association for ecological purposes, y_a and y_b are usually unequal. Therefore, the proportion of quadrats containing species A in which A and B co-occur $[x/(x+y_a)]$ usually does not equal the percentage of quadrats containing species B in which A and B are found $[x/(x+y_b)]$.

Different association indices reconcile these two values in different ways (reviewed in Hubalek 1982). A single index of association can be derived from the two measures of association using the maximum value, the minimum value, the arithmetic mean, the geometric mean, and, perhaps most widely used, a weighted mean known as Sorensen's index or Dice's index (Pielou 1969). The common aspect of all these indices is that they are a reconciliation of two real values known with some confidence, and the index represents no real quantity (hence the use of the term 'index' rather than 'measure' of association).

Table I. Definitions of notation used for estimating the association of two individuals or groups, A and B (after Cairns & Schwager 1987)

x	Number of observation periods during which A and B are observed together
y_a	Number of observation periods during which only A is observed
y_b	Number of observation periods during which only B is observed
y_{ab}	Number of observation periods during which A and B are both observed in separate groups
N	Total number of observation periods
D	Number of observation periods during which neither A nor B are observed
n_a	Total number of observation periods during which A is sighted ($=x+y_a+y_{ab}$)
n_b	Total number of observation periods during which B is sighted ($=x+y_b+y_{ab}$)
T_a	Number of observation periods during which A is sighted without B ($=y_a+y_{ab}$)
T_b	Number of observation periods during which B is sighted without A ($=y_b+y_{ab}$)

Although in ecological studies the association in space of species A with species B does not necessarily equal the association of species B with species A, there exists in behavioural studies a single real value for the amount of time individuals A and B spend together. The values calculated by a behavioural researcher for the percentage of time A spends with B $[x/(x+y_a)]$ and the percentage of time B spends with A $[x/(x+y_b)]$ differ only because of a viewing bias; when A and B are apart, the observer may see one individual more frequently than the other. An accurate measure of association must take into account the transitivity of these two

---

†Present address: Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, U.K.

0003–3472/92/070377+03 $03.00/0

values. Further bias can be introduced to an estimate of association either in data collection or data analysis and are reviewed by Cairns & Schwager (1987).

The problem with the two ecological indices most commonly used by behavioural biologists is that by ignoring this transitivity, they provide inherently biased estimates of association. The most commonly used index is the half-weight index, also known as Dice's or Sorensen's index:

$$\frac{x}{\frac{1}{2}(n_a+n_b)} \text{ or } \frac{x}{x+y_{ab}+\frac{1}{2}(y_a+y_b)}$$

The index is simple to calculate as only three data points are required. Unfortunately, this index tends to overestimate levels of association. By taking an average of the number of sightings y_a and y_b, rather than summing their values, Sorensen's index reduces the value of the denominator and thus increases the estimate of association.

In a similar fashion, the twice-weight index only requires three measures, but leads to underestimates of levels of association:

$$\frac{x}{x+T_a+T_b} \text{ or } \frac{x}{x+2y_{ab}+y_a+y_b}$$

If, in a sample period, A is seen in a group without B, then B must likewise be in a group without A. But the twice-weight index counts each observation as an independent event, 'double counting' all observation periods in which both individuals are seen separately (y_{ab}). This inflates the denominator and decreases the estimated level of association.

Using a computer simulation of patterns of association, Cairns & Schwager (1987, Fig. 2) demonstrated that these inherent biases hold across a large range of observer bias. While within any study, these two indices will provide identical ranking of pairs, comparison of levels of association among studies is inadvisable.

The simple ratio is inherently the most accurate because it neither 'double-counts' sightings in which both individuals are seen separately, nor does it 'deflate' the denominator by giving an 'average' value of sightings of either individual:

$$\frac{x}{x+y_{ab}+y_a+y_b}$$

Unfortunately, the simple ratio is infrequently used (but see Guinness et al. 1979; Dunbar & Dunbar 1981; Clutton-Brock et al. 1982). In large data sets, calculating y_{ab} for each pair of individuals can be time-consuming. We suggest using a simple calculation which is equivalent:

$$\frac{x}{x+y_{ab}+y_a+y_b} \text{ or } \frac{x}{N-D}$$

the number of viewing periods in which A and B are seen together, divided by the total number of viewing periods minus the number of viewing periods in which neither A nor B is sighted.

Cairns & Schwager (1987) show that if certain assumptions are made about biases inherent in data collection, the use of a maximum likelihood estimator can lead to a greatly improved estimate of the time two individuals spend with each other. Despite its clear advantages, the use of a single maximum likelihood estimator for each data set assumes that each dyad is subject to similar sampling errors. Rigorous analyses may thus require multiple maximum likelihood estimators for each study.

Cairns & Schwager (1987) tacitly acknowledge that not all researchers will adopt the use of a maximum likelihood estimator, and this seems to be the case (e.g. Green et al. 1989; White & Burgman 1990). In such cases, they suggest that researchers assess the direction of the bias in their data and select the measure with a bias that counter-balances biases believed to be generated in data collection. We perceive several problems with this suggestion: (1) this suggestion invites 'index shopping'; (2) in many studies, biases will differ among dyads due to differences in ecology and social behaviour (reviewed in Cairns & Schwager 1987); (3) the indices use arbitrary weighting which in no way reflects the extent of bias observed. We suggest that, failing the use of a maximum likelihood estimator, researchers adopt the statistically unbiased simple ratio and discuss how biases might effect results observed. In the end, the judgement of each researcher, accompanied by a full disclosure of rationale, remains the best check on bias in estimating association between individuals or groups, and whether statistical comparisons among studies are appropriate.

The authors would like to thank C. Moss for provoking this note, J. Poole, P. Lee, D. Rubenstein, S. Cairns and S. Schwager for comments on an earlier draft.

REFERENCES

Cairns, S. & Schwager, S. 1987. A comparison of association indices. *Anim. Behav.*, **35,** 1454–1469.

Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. 1982. *Red Deer. Behavior and Ecology of Two Sexes.* Chicago, Illinois: University of Chicago Press.

Dunbar, R. I. M. & Dunbar, E. P. 1981. The grouping behaviour of male Walia ibex with special reference to the rut. *Afr. J. Ecol.*, **19,** 251–263.

Green, W. C. H., Griswold, J. G. & Rothstein, A. 1989. Post-weaning associations among bison mothers and daughters. *Anim. Behav.*, **38,** 847–858.

Guinness, F. A., Hall, M. J. & Cockerill, R. A. 1979. Mother–offspring association in red deer. *Anim. Behav.*, **27,** 536–544.

Hubalek, Z. 1982. Coefficients of association and similarity, based on binary (presence–absence) data: an evaluation. *Biol. Rev.*, **57,** 669–689.

Pielou, E. C. 1969. *An Introduction to Mathematical Ecology.* New York: John Wiley.

White, F. J. & Burgman, M. A. 1990. Social organization of the pygmy chimpanzee (*Pan paniscus*)—multivariate analyses of intracommunity associations. *Am. J. phys. Anthropol.*, **83,** 193–201.