

Darwin a better name than Wallace?

Sir—C. Ambrogi Lorenzini (*Nature* 384, 508; 1996) highlights the possibility that double-barrelled names may be incorrectly listed in publication databases. This is an important issue because citation rates are increasingly used to assess the value of research. But it is not only those with unusual names who may get a raw deal out of citation indices. Investigation of the relationship between alphabetical position of surname and citation rate reveals that researchers nearer to Darwin may appear to be making a bigger contribution than those nearer Wallace.

The number of papers published by authors of a particular surname initial for 1994 was taken from the on-line *Science Citation Index (SCI)*. Citation rates by initial are not easily gleaned from this source, so I measured the number of column centimetres of citations in the paper version of the 1994 *SCI*. Comparison of these data reveals a clear decrease in citations per publication with surname initial (linear regression $r^2 = 0.33$, $F = 12.04$, $P < 0.002$). An obvious

explanation is that authorship of papers is sometimes determined alphabetically. Citations are based only on first-author papers, whereas the on-line database provides information on all authors, so those with later initials may be penalized through being less likely to be first author.

This doesn't, however, appear to be sufficient explanation. By measuring column centimetres per author in the index of 10 randomly chosen 1994 editions of *Current Contents*, which lists only first authors, one is able to estimate publications by first authors only. This also reveals a significant decrease in number of citations with position in the alphabet (linear regression $r^2 = 0.33$, $F = 16.7$, $P < 0.0004$). Because the citation index goes back many years, this result could arise from a past tendency to put authors in alphabetical order. But using order of authorship in Letters to *Nature* from the first issue of the past 25 years fails to support this idea. Taking into account the increased number of authors on more recent papers, there is no decrease

in the proportion of papers with alphabetical authors since 1972 (linear regression $P = 0.17$).

So what is the explanation? Could it be that researchers tend to flick through reference lists and to have had their fill before they reach the end? Do we go to the library with an alphabetical list of papers to read and run out of steam half-way through? Certainly the effect is very robust; removing the five least cited initials from either regression does not drop significance below 0.05, so clearly something is going on.

I suggest that, whatever the explanation, in fairness to those relegated to the end of reference lists, a correction factor should be applied using the slope of the regression of citation rate versus author initial. By this method, the citation rate of someone whose initial is, say, T should be multiplied by 1.36 — quite a boost.

Tom Tregenza

*Department of Biology,
University of Leeds,
Leeds LS2 9JT, UK*