

Personal Paper

Choosing the programme for an international congress

D R APPLETON, D N S KERR

British Medical Journal, 1978, 1, 421-423

Summary and conclusions

The method of selecting abstracts for an international congress at which only 15% of submitted papers could be accepted entailed a panel of 12 assessors using their specialised knowledge, but presentation of the abstract was also important in selection. There was only a limited agreement between assessors in arranging abstracts in order of merit, so that a single assessor would be unacceptable.

Use of the full panel to grade all abstracts was very expensive, but it could be replaced, without unacceptable injustice, by dividing the work randomly among groups of three selectors.

Introduction

International congresses are often reviled and ridiculed. Some maintain that they exist solely for the benefit of the travel industry, but, as they continue to grow and proliferate, they probably meet some deeply felt need of scientists as well. At many the number of abstracts submitted for presentation grows from year to year; the programme is usually selected in one of two ways.

Selection procedures

The all-embracing approach—Plenary sessions are held in the morning with invited speakers. Submitted original papers are relegated to parallel sessions in the afternoons. Many delegates decide that if you are missing 13 sessions you might as well miss 14 and see the Taj Mahal/Uffizi Gallery/Golden Gate instead, so that these parallel sessions may be held in small seminar rooms, of which the modern conference centre has an almost limitless supply. Consequently there is little restriction on the number of abstracts accepted, and a small selection committee can quickly weed out the rubbish and divide the rest into convenient packets that fill the afternoons. This is tough on the last speaker, who addresses his wife and the previous speaker (if polite), but it ensures that most delegates get their travelling expenses, which allows the organisers to plan ahead and book attractive (and expensive) speakers for the mornings. This approach calls for good business acumen but presents few problems to the selection committee.

The selective approach—Submitted papers are given pride of place. They are delivered in plenary session and published. Their number is limited by the length of the meeting and the capacity of the unboggled human mind. Competition for a place on the podium is keen, its achievement an accolade, and in some societies a condition of full membership. Selection of papers must be fair and must produce the best programme possible.

We favour this approach but recognise its difficulties. It can work only if busy professionals, in their limited spare time, can read an abstract among several hundred and make a reliable judgment on it. Their task is made easier by the strict rules that are now enforced: abstracts must give explicit results and conclusions; promises that "the results will be discussed" are inadmissible; and they must be clearly typed in a stipulated space. None the less, an abstract that may have taken hours to write is read and judged in a couple of minutes. In an effort to increase reliability some societies use many assessors and pool their results but this is expensive in manpower. We have therefore analysed the selection exercise of a society using the selective approach to see how far it combines reliability with economy of effort.

Methods

The European Dialysis and Transplant Association (EDTA) holds an annual congress devoted to dialysis, transplantation, and general nephrology. The subdivision of time between these subjects is decided in advance. For the meeting at Helsinki in 1977, 329 abstracts were submitted (317 in time for this study) and 50 papers accepted. Authors were asked not to identify themselves in the text and to send additional photostats of their abstract with the authors' names and institution obliterated. These anonymous abstracts were sent to 12 selectors who were members of the elected council and whose clinical and research interests spanned those of the society. Each marked abstracts on a six point scale from "0=reject" to "5=must accept" on the basis of scientific merit (defined as originality, adequate data, legitimate conclusions, and interest to the society). Selectors did not mark abstracts from their own institutions and for these the average of the other 11 marks was substituted in the selection process; the aggregate mark determined the paper's chance of acceptance until the appropriate section of the programme was filled. Nevertheless, for this statistical exercise most selectors kindly gave us confidential assessments of the papers from their own units, which we substituted for the average mark. In a few cases a selector did not mark an abstract, presumably because he felt unable to judge its contents.

To test the hypothesis that the presentation of an abstract is more important than its content, we submitted a duplicate set to a medical statistician with no specialised knowledge of nephrology (DRA) who marked them on presentation and general scientific approach. The 13 sets of marks were then analysed.

Results and discussion

HOW VARIABLE ARE SELECTORS?

Selectors are disappointingly variable, but not surprisingly so. The number of papers rejected varied from none to 23 and those marked

Departments of Medical Statistics and Medicine, University of Newcastle upon Tyne, Newcastle upon Tyne

D R APPLETON, PHD, lecturer in medical statistics
D N S KERR, MSC, FRCP, professor of medicine

"must accept" from none to 13. The mean mark varied from 1.85 to 3.12. This would not matter if selectors simply varied in the severity of their marking but arranged abstracts in much the same order over a similar scale.

DO THEY MARK OVER THE SAME SCALE?

No, the marks are not over the same scale, but the differences are less important than might be expected. The standard deviation of their marks varied from 0.83 to 1.22. Ideally the marks should be scaled to the same mean and standard deviation before aggregation but doing this altered only two of the top 25 and two of the top 50 papers in this study. Scaling the marks would present no problems to most university centres if they were received well before the meeting of the selection committee, but in real life this is unlikely to happen.

DO THEY ARRANGE ABSTRACTS IN THE SAME ORDER?

No, the abstracts are arranged differently. Seven abstracts received 0 from one or more selectors and 5 from another. There is, however, a significant correlation among the marks given by different selectors. We compared the marks of each official selector with the mean of the other 11 selectors for 288 abstracts on which we had a complete set of marks. The squares of the correlation coefficients (roughly the proportion of agreement) varied between $r^2=0.05$ and $r^2=0.29$. For this number of abstracts any value of r^2 over 0.023 is significantly different from zero at the 1% level.

Each of the selectors was compared with all the others. The closest correlation ($r=0.41$) was between two physicians of the same nationality, background, and research interests. Not all the correlation coefficients were significantly different from zero, and the average was $r=0.20$. This may be compared with the work of Matthews and Leece¹ on impression marking—also on a 0-5 scale—of questions in A-level chemistry examinations; there the average correlation coefficient was 0.76. There was, however, an opportunity for bias in the latter study that did not exist in our own, and better agreement among examiners judging answers to the same question would be expected than among those assessing the relative merit of diverse research projects in a wide scientific field. The rather low correlation coefficients in this study probably represent what is achieved in programme selection and are a warning that the use of one single examiner, though commonplace in examination marking, is unlikely to prove suitable for selecting abstracts.

ARE PAPERS PICKED ON CONTENT OR PRESENTATION?

A cluster analysis (single linkage, using 1-r as the "distance" between two assessors) was performed, and the resulting dendrogram is shown in the figure. Such a structure is a pictorial representation of the clusters being formed: those assessors who are most similar (5 and 6) join together first, then number 4 joins them, numbers 10 and 11 then form a separate group, and so on. There comes a point when the distance between individuals in different clusters is large enough for us to consider the clustering process to be finished. It would seem reasonable in this study to say that this occurs in the gap before the group containing assessors 1 and 2 joins the main group. The dendrogram can therefore be interpreted as showing that 10 of the official selectors tended to group together, the other two being rather dissimilar from one another as well as from their colleagues. This may reflect their respective interest in two of the topics that are of deep concern only to a minority of the society, as reflected in the abstracts submitted.

We were gratified to find that all the official selectors clustered better with their colleagues than did the statistician, so they are presumably using their special knowledge in making their judgments. The correlation ($r=0.17$) between the statistician's marks and the mean of the others was significant at the 1% level, however, suggesting that format and general scientific approach are important factors in selection.

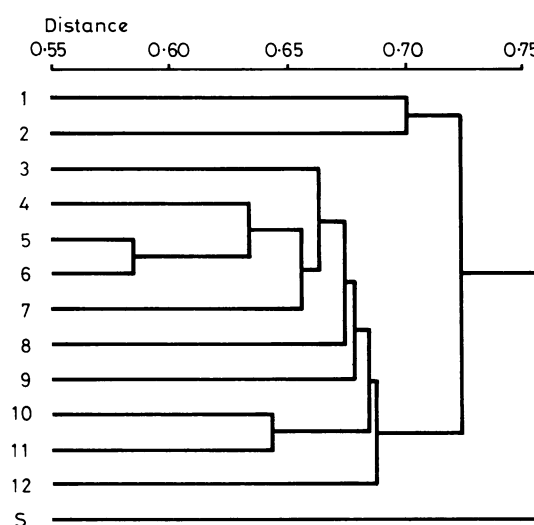
The widespread use of non-standard abbreviations for items of equipment, chemical compounds, drugs, operations, and diseases hindered our appreciation of the abstracts. Explaining the abbreviation the first time it is used does not make the abstract any more readable, though it may make it comprehensible with effort. The banning of all non-standard abbreviations would reduce by a small amount the

data that could be presented in an abstract (no bad thing, we reckon) but would more than compensate by making it much more digestible by selector and participant alike.

COULD THE PROGRAMME BE PICKED MORE ECONOMICALLY?

Twelve physicians and surgeons, with an average age in the 40s, read these abstracts. Extrapolation of our own reading time suggests that they devoted at least 200 man-hours to the task in aggregate; this may be an underestimate, since most of the assessors, though fluent in English, had it as a second language. The time of these selectors, who are bound to be prominent members of the profession, is expensive and limited. Any further increase in the number of abstracts could easily overload the system.

We have therefore calculated what difference it would have made to the programme had fewer selectors been employed or had each marked only some of the abstracts. We shall refer to the top 25 and 50 papers picked by the whole 12 selectors as "the best" 25 and 50.



Dendrogram for single linkage cluster analysis using 1-r as distance function. Official assessors are indicated by numbers 1-12 and the statistician by S. For explanation see text.

We have compared with them the selection that would have been made by all permutations of six, or of three, selectors. We aggregated the marks of these subgroups to see how their top 25 and 50 compared with the official best 25 and 50.

The most successful (or representative) group of six would have chosen 45 of the best 50, and the most successful three would have chosen 40 of the best 50. The programme could have been picked much more economically, with little injustice, if we could have chosen three selectors with foreknowledge. Inspection of the most successful groupings in our study, however, has not suggested ways in which they could have been picked in advance. They could therefore only be assigned this dubious honour in second and subsequent years, and we doubt whether there would be many volunteers to read all abstracts year after year. Moreover, we do not yet know whether performance is consistent from year to year.

It is therefore of greater interest to know what would have transpired with random selection of assessors. Groups of six selectors would have picked anything from 27 to 45 (mean 38) of the best 50 and an average of 19 of the best 25. Groups of three selectors would have chosen an average of 33 (SD 3.6) of the best 50 and 15 of the best 25.

The method most likely to appeal to selectors is the distribution of each abstract to a random selection of three out of the 12 assessors, so that each has an equal quarter share of the total work. Simulation studies show that this procedure leads to a mean of 32 (SD 2.5) of the best 50 and 14 of the best 25 being picked. This is about the same result as we obtained by asking three selectors to do the whole job but the outcome is a little more predictable, as shown by the smaller standard deviation.

Comment

This last method, which we commend as equitable and practicable, even in the face of an avalanche of abstracts, contains some rough justice. An international congress is expensive. The Helsinki meeting was attended by over 1500 delegates from as far afield as Australia, Japan, and Western America. Even those travelling from Newcastle, using group travel and cheap accommodation, spent £230-330. The selection of a good programme is therefore important and worthy of the effort now expended on it. A well-tryed method should not be lightly discarded. Nevertheless, the combined opinions of 12 selectors can only approximate to the choice of all participants.

Most of the "errors" that would have been committed by spreading the load as we suggest represent displacement of papers from just above the cut-off point to just below or vice-versa. When the problem is an embarrassing superfluity of good abstracts this cannot be disastrous. The best 50 abstracts do not necessarily lead to the best 50 papers, and the marking process

is only the first step in constructing a balanced programme. We therefore feel that the modest measure of injustice inherent in our proposed scheme is an acceptable price to pay for speedier and less exhausting work by the selection committee. A late deadline for submitting abstracts and an early announcement of the programme are other benefits that would accrue from spreading the selection load.

We are grateful to the president of EDTA, Professor Jørn Hess Thaysen, and the members of its council for permission to carry out and publish this study and for their helpful comments on the draft manuscript.

Reference

¹ Matthews, J C, and Leece, J R, *School Science Review*, 1975, **57**, 362.

(Accepted 13 December 1977)

Letter from . . . Canada

The new Sinbad syndrome

PETER J BANKS

British Medical Journal, 1978, **1**, 423-424

Those of us that go through the flaming hoops of medical officialdom learn to listen with great care to those whose primary skill is in communication. These men, whose unhappy, but invaluable, task it is to shepherd us through the labyrinthine traps that await the unwary when dealing with the media, tend to be large, cautious men of equally solid opinion. It is always an opinion well worth attention.

When, therefore, a letter winged its way from Ottawa, bearing the stamp of Mr Douglas Geekie, our National Director of Communications—that is, if any letter can be said to wing its way anywhere through the contemporary Canadian postal system—I gave it my closest attention. My good friend, in his usual forthright manner, suggested that in some of my recent writings I had begun to show a note of antigovernment paranoia that was not becoming. Now, some of my colleagues from Ontario and the eastern prairies are well known for possessing a degree of political sophistication, at least as well developed as that of the hunter-gatherers of the pleistocene era, but in the two most westerly provinces we have always felt that the development of a working relationship or a partnership with government is more than a political slogan. Was this merely a delusion of western euphoria, or has there truly been a change in attitude?

There is, of course, always a basic difference between those who administer the healing arts and those who perform them. The administrator hesitates to make an exception for fear that, should the exceptions multiply, administration would degenerate

into chaos. To the doctor every case is a potential exception to be judged and treated on an individual basis. It is often difficult to bridge this essential difference in outlook, but that has always been with us, as it has always been with business men, farmers, fishermen, and all those who try to go it alone. Yet, across Canada, there is growing a very definite change in attitude towards government, and I am not referring to the age-old and well-justified distrust of politicians but to the very definite change in the average Canadian towards what used to be called "the Civil Service."

Non-productive tasks

In the past 50 years, as Western democracies under the constant spur of universal suffrage have become dedicated to achieving the biological absurdity of human equality, so the bureaucracy, whose job it is to attempt to bring this impossibility to reality, has multiplied in their numbers and in their areas of activity. The phenomenon is now quite beyond the control of elected government. Politician after politician that the long-suffering taxpayers have elected on the platform of reduced government expenditure find themselves powerless castrates in the hands of their non-elected mandarins. More and more of our national resources are channelled into the non-productive tasks of regulation and control and confiscation. The people of the Western democracies now find themselves supporting a huge army more burdensome than even that carried by our forebears, whose gilded monarchs paraded their pride in shining battalions. Inevitably, as always in history, when a society multiplies the numbers of the non-productive, whether in a priesthood or in an army or in an administrative bureaucracy, then the debasement of the currency or inflation is the inevitable consequence.

Victoria, British Columbia

PETER J BANKS, MD, FRCP, consulting physician
