

THE RATING COMMITTEE REPORT: A PANEL STATEMENT
Raymond J. Jessen*, C-E-I-R, Inc.

Fortunately, or perhaps unfortunately, much of what I have to say on Broadcast Ratings is contained, together with those of my colleagues, in the report of our committee which was submitted to the American Statistical Association last January and published by the U. S. Congress as House Report #193. Today, in order to help get a discussion going for this panel I shall make some remarks bearing on the following 4 points:

1. A brief summary of the complaints on the broadcast ratings and our reaction to them
2. The Public Interest versus the Advertiser Interest in Ratings
3. Requiring clear Labelling versus setting Standards
4. Need for methodological Research

I. Summary of Complaints on Ratings and our Reactions to them.

Sifting through the complaints that one hears about the ratings the following three rather broad reasons seem to have some bearing on how or why the complaint may have arisen.

- (1) The ratings are not really what they pretend to be,
- (2) The ratings are what they pretend to be, but are inaccurate and
- (3) The ratings should be something more than what they pretend to be.

In the first group the complaints appear to be concerned with the matter of coverage, both that of time and space. Some rating services use or have used a specified period such as a week, a/ for example, during which their data were collected but in the report may imply or declare that the rating applies to some longer period of time such as a month or a quarter. If a broadcaster has exclusive knowledge of when a rating period is scheduled he has opportunity to offer special attractions during that period (known as "hypoing") to make a better showing than he would otherwise. But even if all broadcasters know when the rating period occurs then disputes may arise as to the "representativeness" of whatever period is selected. However this practice of basing ratings on short periods when longer periods are presumed to be represented appears to be used much less now than several years ago - and therefore this problem is disappearing.

In addition to the complaints which arise because of restricting the time period covered by ratings, are those that arise because of restrictions on the area or the segments of the population covered. Some ratings for example fail to include such things as: certain outlying

portions of a market or broadcast area, persons not-at-home at the time interviewers call, certain foreign speaking homes etc. Usually these omissions are for economic reasons but the usefulness of the results is so seriously impaired thereby that the economy may well be false.

Several of the services could improve the understanding of the appropriate period of time and to what universe for which their results were applicable. The use of a "survey week" to represent the average goings-on during a month not only leads to fruitless arguments on how well it represents some longer time period but is wasteful from a sampling point of view. Confusion on just what universe is being dealt with is also not difficult to eliminate if the service will clearly present it in his report. However the effects of the omission of homes because they are too difficult to reach because of distance, language, etc., is difficult to assess without considerably more evidence than is now available.

The segments of homes omitted when relevant by nearly all the services are those in which a respondent is not-at-home at time of call, those not listed in a phone book, and those which refuse to cooperate. Regardless of method used, these groups appear to average about 40% of all eligible homes - a rather large group on which to have no viewing or listening information. In interview surveys, this group could be reduced substantially if call-backs were made; but in diary and surveys the non cooperators are a difficult problem and no direct approach seems feasible.

The second group of complaints seem to arise not because the ratings are erroneously conceived but because they are inaccurate. Examples of these complaints are: "hypoing" (where calculated misdeeds are attributed to the broadcaster), fraudulent results (where calculated misdeeds are attributed to the rating service), technical faults and weaknesses attributable to the ignorance or carelessness of the rating services and to sampling variation, which might be blamed on the fates as well as on the rating services. Included in this class are such charges as, "small samples cannot properly represent the diversity of human behavior in this county", "scientific sampling may be all right for determining such things as the number of refrigerators or garden tractors but dealing with people is another matter", "I know of interviewers who just sit in their homes and fill in the answers!", "Rater Q gave a rating of 10.2 to a program that didn't exist!", etc.

Even if its purpose is crystal clear, there are a number of inaccuracies in almost any survey. These inaccuracies or deviations from the true are generally classified by survey statisticians into the two types: (1) sampling deviations and

*Member of the Technical Committee on Broadcast Ratings

a/ Trendex uses 1 week per month. Pulse dropped the single week period about three years ago. Hooper uses 1st week in month. Sindlinger is continuous.

(ii) non-sampling deviations. The likely sizes of deviations due to sampling can be determined if the type of sample is appropriate (such as random, for example) and its size is known. Some of the rating services provide the means for making this determination, some don't. In the case where clusters are used at some stage in the sampling some of the services fail to provide the user with information on how this feature affects sampling precision. There are some instances when the effect of clusters must be quite large and therefore seriously misleading if only the number of homes sampled is given.

Non-sampling deviations arise from errors in the original data, failure to obtain data from all designated homes in the sample, processing errors, errors inherent in the method of measurement, etc. These deviations in ratings are a more serious problem than those of sampling because they can not be made determinable through correct statistical design and generally require special studies for their detection and measurement. Although the raters presented some studies on this problem, they were of little help for judging the effects of such deviations on the accuracy of the ratings.

In the third group the complaints arise because the complainant seems to expect something more than that which the trade ordinarily understands to be a rating. The complainant takes a dim view of the utility of "nose-counts" and demands ratings that measure the "quality" of the audience, the satisfactions it receives, its advertising or commercial effectiveness, etc.; or perhaps the program's effect on the morals, culture and well-being of the people so that "public interest" can be more adequately evaluated. Also it is felt by some that the needs and desires of those not in the audience might also be included in some kind of "rating".

I am inclined to be sympathetic with these complaints, and agree that there is a need for statistical measures other than "nose-counts" so we can determine the nature and feelings of the audience as well as its size. However the ratings in their present form can and do serve useful functions. But this does not mean we should not be looking for something more.

II. The Public versus the Advertiser Interest in Ratings

There is a basic difference in the needs for ratings among the different interest groups in and out of the broadcast industry. Advertisers are mostly interested in some simple measure of the effectiveness of broadcasting as a means for selling their products. A program is a good program and a station or a network is a good station or a good network if it does an efficient job of stimulating the advertiser's sales. Since advertisers usually use a number of different media at the same time and since the advertising effectiveness of any particular campaign may be indirect as well as direct, it is very difficult to determine the sales effects of any broadcast program; or even for any part of the firm's overall

advertising program. In the absence of these unavailable measures of effectiveness the advertiser depends on the use of audience measures such as size, type, etc. - measures of characteristics which he believes are related somehow to his actual or potential sales. Thus broadcast ratings give the advertiser a means for evaluating the expected effectiveness of television, newspapers, magazines, etc. for his advertising dollar. They may not be very good for this purpose but they are available. When the advertiser feels the ratings are not doing a good job of measuring advertising effectiveness he may be inclined to complain about their inaccuracies and other inadequacies.

On the other hand there is a group which has little interest in the matter of advertising effectiveness but rather is concerned with what it regards as the "public interest". People in this group are generally interested in how "moral", or "educational" or "cultural" a program might be. Obviously the standard ratings do not measure these things, nor is it very obvious how they can be measured. I personally feel that the problem of finding suitable measures should not be ignored and I would like to see the behavioral and social scientists devote more effort to its solution.

III. Standards versus Labelling.

A fixed set of standards for ratings has a strong appeal to many people in the industry - particularly those who must make many decisions using many different ratings from the different rating services. The simplicity and economy of some industry standards would certainly reduce the burden of evaluating the different ratings as to accuracy and comparability and facilitate communication and understanding among the various persons involved. For example, if everyone used the same "adjustment factor" to translate rating X into rating Y a lot of time and argument might be saved. The industry can do much on its own to bring about some standardization, and this would be facilitated if an industry-wide office of methodology is established.

But there are certain kinds of standards that I feel ought to be avoided, such as for example, setting standard sample size requirements. Since samples of different designs may differ in precision for a given number of homes contacted, a size requirement in terms of a fixed number of homes may encourage the use of inefficient sample designs. Setting size requirements in terms of random sample equivalence may avoid this difficulty, but of course at a cost of some loss in simplicity and understanding of what is happening.

On the other hand if certain accuracy requirements are established, rather than sample size, then there is a danger that they will be too high for some purposes and perhaps too low for others. The rating business is fairly technical and will perhaps become much more so, but it seems inevitable that many concepts and practices in the rating business will become more

standardized, I hope that paralleling this development will be an increased use of technically trained people to cope with the more sophisticated, more useful and more efficient concepts which also seem almost sure to come.

In order to avoid some of the undesirable "side effects" of standards, I suggest a more vigorous attempt to adopt the practice of "clear labelling". By this I mean that if rating services would clearly state what their ratings mean statistically, such as the proper projectable universe, the sampling error, the non-sampling error (as well as it can be determined) and of course a statement somewhere of the procedures used, we would have less pressure for standards. By labelling fully and clearly, good practices may have a fighting chance to win over bad practices. And new and more useful approaches will have a better opportunity to find a possible market. Also technical development and technical understanding would be speeded in an industry where progress has been too slow.

IV. Need for Methodological Research.

In the earlier parts of this statement I tried to elaborate on the idea that many of the criticisms of the ratings arise because of lack of understanding of one sort or another. Some are due to failure to understand what the ratings are purported to be. Others are due to a lack of understanding of what the various errors in rat-

ings might be in the practical problem at hand rather than in some very specific and perhaps dramatic instance. Still other criticisms are really appeals for ratings which are fundamentally different than the current ones and perhaps would have only a limited market. And of course a lot of criticism is essentially warfare between the protagonist and antagonist of the different methods used and the houses that produce them.

An industry-wide Office of Methodological Research would do several desirable things. It would be an excellent place to bring about an understanding of the underlying common elements in the rating business, which may be overlooked otherwise. This it could do by what would be essentially an educational program. By preparing and issuing reports on procedures used by each service and the supportable declarations that each gives on the advantages of his procedures. In addition - and this would be the primary function of the office - it would itself carry out or would sponsor research in problems common to the industry. These might be problems of setting standards, of clear labelling, of determining non-sampling errors, of determining suitable concepts for measurement, etc. It would be used otherwise, but I hope it would be used to encourage progress in a very important industry - what with the coming of the computer age and its appetite for more and more data - an industry which will become much more important in the future.