

Are the peplemeters valid? The Spanish experience.

Carlos Lamas - AIMC, Madrid.

The first peplemeters in Spain were installed in 1986 but, although some experimental results were presented in 1987, the measurement system was not fully operational until the middle of 1988. The start was a tender opened by the state-owned television network (Radio Television Espanola). The meter was “Spanish made”, being designed and manufactured by a public company. The “not-invented-here syndrome” and an alleged “support to the national industry” were the reasons to reject the other foreign hardware alternatives. Furthermore, 60% of the stock of the audience measurement company belonged to the Spanish PTT (Compania Telefonica) which is controlled and also partially owned by the Government.

But this remarkable state involvement in the TV audience measurement operation began to disappear gradually. The whole market (TV stations -both public and commercial- and the advertising sector) started to use the meter ratings as the official currency to sell and buy commercial airtime. The Swiss Telecontrol meter replaced the big sized and failure-prone Spanish box and today the operation is principally owned by the well-known French group Sofres (a private independent firm).

I will not go into the details of the troubled history of the meters in my country, but I cannot resist the temptation of mentioning the scandal of the publication of the names and addresses of all the panel members in January 1991. The list was published in the Sunday issue of one of the most important national newspapers. As far as I know, that was the only case worldwide in which the anonymity of the members of an operational panel was broken to such a significant extent. The disclosure of the names of the panel members was due to a leakage by a disaffected former employee. The reasons behind the newspaper action were never clearly justified and, as you can imagine, all the panel members had to be replaced with a considerable related cost.

That story brought one positive element to the control of the meter operation. A Users Committee was set up to monitor the actions being taken to re-establish the validity of the panel. Furthermore, that Committee was given the responsibility to advise, supervise and eventually decide on all the methodological matters related to the meter panel operation. This group played a very important role in the development and improvement of the Spanish audience measurement system.

The controlling activities performed by the Users Committee were complemented and extended when the companies and institutions involved in the measurement of television audience ratings -namely television stations, advertisers, media buying groups and advertising agencies- commissioned AIMC to assume, on behalf of the industry, the supervision and auditing of the measurement operation. The idea was to support the principle of a single audience rating system subject however to a detailed and in-depth control. It was thought that this type of control needed a dedication and expertise which were beyond the possibilities of the User Committee members. Furthermore, that control required access to sensitive and confidential inside information from Sofres AM (panellists names and addresses for instance) which must necessarily be very restricted.

AIMC, Asociacion para la Investigacion de los Medios de Comunicacion (Association for Media Research) is the Spanish JIC body made up by representatives from the media, advertisers and companies in the advertising sector. One of its activities is to carry out the Estudio General de Medios (EGM), a multimedia research project which provides the audience figures used by the Spanish market for the radio and the press.

The aims set up by the AIMC when taking over this audit function were, in very broad terms, the following:

- Assess from a technical standpoint the different elements of the audience metering operation.
- Check compliance with agreed operating and quality standards.
- Promote better quality information on television audiences.
- Give technical advice to the research users (represented by the Users Committee).

To give an overview of the activities performed, we can schematically classify them into the following areas :

Universe determination : Continuous updating of the size and characteristics of the population to be measured. Establishment survey. Coverage / place of viewing analysis : second homes, out-of-home viewing (at the place of work, at bars or restaurants, at school, etc.), collective housing (army barracks, hospitals, convents, hotels, old peoples' homes, etc.), etc.

Sample : Sampling design, household selection procedures, control of representativeness according to the most relevant variables and descriptive analysis on different panel behavioural and functional indicators (sample turnover, geographical dispersion, stability along time, panel seniority, meter hardware failures, etc.).

We also perform periodical checks on the panel households to verify that they actually exist and to validate their registered characteristics. These checks, done on a selected sample of the panel, are made through telephone calls.

Collaboration bias : In order to study any possible bias the panel might have in relation to total television viewing, an analysis was made using the 1991 establishment survey - which was used as frame for the selection of the new panel- comparing the declared television viewing time per day among the households that joined the panel and the overall total. Above-average television viewing in panel households was detected, confirming the existence of a bias towards over-estimating audiences in the meter estimates.

Meters : We only performed simple tests under laboratory conditions, to confirm the meters produce the expected results when a list of actions is carried out on the TV set or the VCR.

Data Processing : Checks on programs and procedures in connection with :

- Weighting factors calculation
- Conversion of the meter statements into a minute-by-minute and individual-by-individual data base.
- Data aggregation and data projection.
- Audience indicators calculation.

In this area the approach has always been to perform a parallel processing using programs developed by us and compare our results with those obtained by Sofrés AM. The good quality of a research can easily be offset by the errors in the production software. There is no such thing as an error-free software system. And when there is a mistake in the programs, its effect is usually not random but goes in a systematic direction producing a bias in the final results. Therefore, we have always considered that this area deserves a special effort. Through our checks, we have found out quite a number of program bugs which were promptly corrected by the meter company. And although in those cases the incidence on the audience estimates was not serious, the very existence of errors underlines the potential risk.

Other quality checks

- Visits to the panel households. (Correspondence between tuning voltage and TV stations between buttons and individuals).
- Internal coincidental studies.

Instead of directly acknowledging channels, the meter records tuning voltage or radioelectric channels tuned in (depending on the particular kind of meter) and the system converts these into television channels following the relevant conversion tables. The tuning voltage corresponding to every TV channel is different from one household to the other and, even within a household, it can change from one device to the other. The reliability of these correspondence tables is thus essential in order to allocate audiences correctly to different television channels. The AIMC carries out three tests per year. In this respect, we select several households within a province, visit them together with technical staff members from Sofrés AM and then compare the information which Sofrés AM used in this regard for the daily process with the real situation measured in the home. During those visits, a parallel check of the remote control buttons allocated to different members of the household is also performed.

The AIMC supervised the two Internal Coincidental Studies carried out in 1993 and 1994 with the basic aim of measuring the discipline level of the panel members with regard to their obligation to identify themselves to the meter when they start to watch television and to sign off when they finish watching. Essentially the idea is to determine by means of a telephone call if a panel member is watching television at a given moment in time and contrast this statement subsequently with the information gathered by the meter for the same moment. The results achieved are in line with those obtained

previously in Spain and also with those from other countries. The latest study, made in October last year, includes the following data taken from a sample of 7714 panel members :

		Meter		
		Not watching	Watching	Total
Coincidental	Not watching	61,4 %	4,8 %	66,2 %
	Watching	5,6 %	28,2 %	33,8 %
	Total	67,0 %	33,0 %	100,0 %

Full coincidence: 89,6 %

Meter / coincidental ratio: 97,8%

There is one interesting methodological aspect in this survey. In half of the sample, the interviewer identified himself as a Sofrés AM employee and justified the call as a check element to verify the proper functioning of the meter. In the other half of the panel, the interviewer said he was calling from a polling company to make a survey on the usage level of different media (press, radio and TV) and the questionnaire was slightly different from the other, to include some simple questions related to the press and the radio. The objective of all this was to have an idea of the potential effect on the panel members of the feeling of being controlled, which might lead to declare as viewers precisely those household members that in the moment of the call were signed on in the meter.

The results for both parts of the sample were the following:

	First half (Sofrés AM)	Second half (Other research institute)
Full coincidence	92,4 %	86,7 %
Meter / coincidental ratio	100,7 % ³⁷	94,8 %

As expected, the first half of the sample achieves a higher coincidence degree, thus confirming somewhat the existence of a “feeling controlled” factor, but the importance of this effect is not extremely high.

The internal coincidental is usually conducted by the same meter company. The justification for this is the safeguard of the confidentiality of the panel members. We accepted that approach but supervised the fieldwork by the following system:

- Every day, a paper copy of all the questionnaires corresponding to the interviews made that day, must be delivered to AIMC around midnight. At that moment, there is no possibility to know the audience statements provided by the meters.
- Also every day, but some hours later, we received the meter-based audience data corresponding to the households being interviewed the day before.

- When the fieldwork was completed, we were given a computer file with the result of the cross-check between the two sources made by Sofres for each and every panel member together with the final report containing the results of the analyses made.
- We then performed the following checks :
 - On a sample basis, that the coincidence status on that file is correct.
 - That the results shown on the report are actually obtained from that file.

Overall validity checks / Comparison with other sources

- External coincidental surveys
- Comparison with the EGM figures (continuous research based on recall)

The first external coincidental with the purpose of “validating the meters” results was carried out in 1991 due to the pressure of the market. The panel had lost its anonymity and, during the long year it took to replace it, was still producing the official TV audience figures. The market needed to restore its confidence in a “damaged” system in order to keep the audience currency in use. Three external coincidental surveys were carried out during 1991, and since 1993 there is one made every year.

As you all know, this kind of study is carried out by an independent research institute commissioned to that effect. It is made by telephone calls to a newly selected sample and the basic question to the respondent household is whether any member is watching TV on that precise instant and, if that is the case, which programme and channel is being watched. It is generally accepted that, if properly done, the coincidental telephone procedure can be regarded as a criterion for evaluating the reliability of any other methodology. It has to be said, however, that the survey is restricted to telephone households - in Spain, the phone penetration is only 83 %-, that it is also subject to some degree of rejections to collaborate and that the treatment of non-answering households is a particularly costly and delicate operation.

The between-methods comparison is usually focused on two parameters :

- Level of Total TV viewing (average rating or viewing minutes per capita).
- Channel shares.

What have been the main findings of that comparison? How do the figures coming from both methods compare?

Let us first talk of the total television viewing. According to the first coincidental survey conducted in 1991, the meters were over-reporting the TV consumption by 0,3% but the second one (carried out two months later and the panel remaining basically stable) indicated that the size of the over-estimation was around 10%. As that was apparently due to the differences in the handling by the two different research companies of the households not answering the telephone call, it was decided to drop the total television viewing measurement from the objectives of subsequent surveys, on the basis of the following considerations :

- The main concern of the users at that time was really not the reliability of the total TV estimates but the relative importance of the various channels.
- By concentrating in estimating the shares, we increased significantly the sampling efficiency. Only households answering the call and declaring that at least one member was watching TV were considered. In other words, for the same household sample size, the number of individuals watching TV (i.e., the relevant sample size for the precision of the share estimates) was considerably higher. And, in order to consolidate the different half hour periods, each one was projected to the total TV viewers estimates provided by the meter panel.
- We were able to avoid the cumbersome procedures for repeating the call to the non-answering households and the debatable question of making assumptions when the household is never contacted.
- As a result, the system became simpler and cleaner.

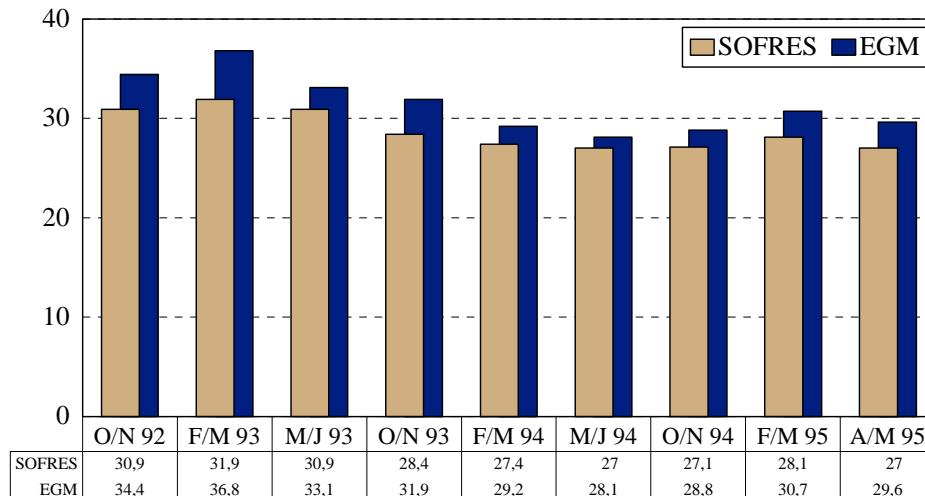
And what does the share comparison look like?

	Nov 93			Nov 94		
	Meters	Coinc.	<i>Differ.</i>	Meters	Coinc.	<i>Differ.</i>
TVE1	30,2	29,9	0,3	30,1	30,6	-0,5
TVE2	9,2	9,2	0	8,6	7,2	1,4
ANTENA 3	24,4	24,2	0,2	24,5	24,7	-0,2
TELE 5	16,8	17,9	-1,1	18	19,5	-1,5
CANAL PLUS	2,5	2,6	-0,1	2,4	2,4	0
AUTONÓMICAS	16,5	15,7	0,8	15,8	14,7	1,1
OTHERS	0,4	0,5	-0,1	0,6	0,9	-0,3

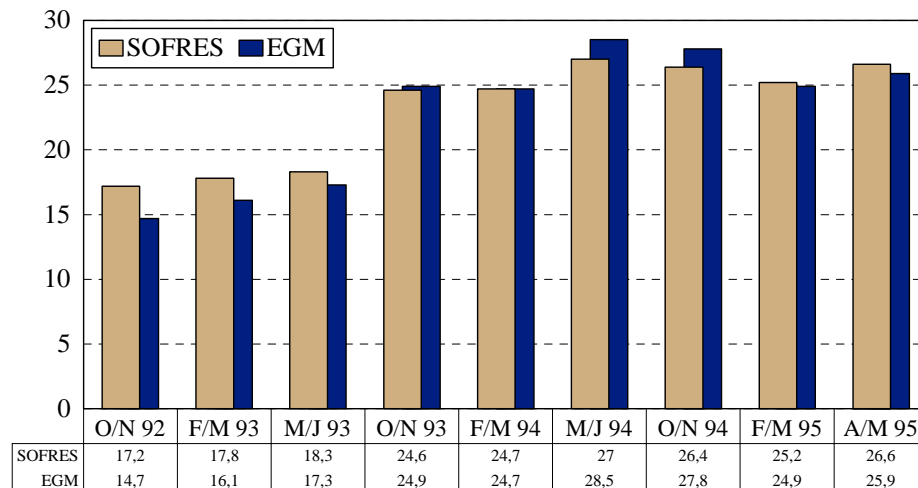
Although most of the differences can be attributed to the sampling variation, there are others that are significant at the 95% confidence level.

The main goal when comparing the audience meter figures with those coming from EGM (recall method) is to check the consistency of the trends. Up to a certain degree, you have to accept some differences in the level of audience estimates, but the trend should be the same. We use the trend of the channel shares for controlling purposes. Let us have a look at the evolution of two channels :

TVE1



ANTENA 3



Accuracy of audience estimates

- Loss of accuracy due to weighting.
- Loss of accuracy due to the cluster effect (people grouped in households)
- Gain obtained through aggregation of minutes.

By calculating the standard errors for thousands of rating estimates and comparing them with the results obtained, making the assumption that we deal with a proportion estimate in a simple random sample scheme, AIMC has tried to quantify the effects of the three factors mentioned above.

The above findings have been applied to develop a short-cut formula to estimate the standard errors of the audience figures provided by the meter panel. In order to promote its usage, AIMC has written and distributed a related Windows program among the users.

Other activities

- Interviews to ex-panellists.
- Analysis of the panel fatigue effects.

By the end of 1994, a qualitative study was made working on people that had been members of the meter panel. Three types of exercises were made:

- In-depth interviews to individual ex-panellists.
- Interviews to family groups.
- Focus groups among ex-panellists from different households.

The objective was to know more about the opinions, feelings and experience of the people in relation to their previous membership on the panel.

AIMC also made an analysis of the effect of panel members' seniority on the reported viewing times, following the model used by CONTAM in USA in 1989.

Evaluation of the methodology

Now it is time to go back to the question which gives the title to my paper : Are the people meters valid? What do I feel in this respect after 3 years of supervisory work? I advance that my answer is "yes" as an answer, but I will try to explain the reasons behind my answer and also mention my reservations and the limitations I find in the methodology.

On one hand, and because of a number of reasons, the research in the field of social sciences cannot be regarded as an exact science. You can never count on very precise tools. Human beings are too complex to be subjected to reliable measurements by a single straightforward system. Therefore, I believe that a method in this field should be judged in comparison to other possible existing alternatives and not on the basis of how far away it is from perfection or from that ideal solution which is never achieved. Taking this into consideration, I believe that a properly handled meter operation can provide a valid, although not perfect, audience measurement system.

Leaving aside the problems the meter hardware can encounter, today or in the future, in identifying and measuring the channels, the most critical and controversial points in the general methodology remain the following :

- **The low response rate** (below 30%). The panel can be hardly regarded as a probabilistic sample and, from an academic point of view, this means that there is no way to quantify the degree of precision in the reported audience estimations. Of course, you have to consider that the same problem also affects to other cases of panel research, for instance the retail panels, and experience has shown that, in spite of the problems of collaboration, retail panels are a valid and precise sales tracking

tool for a number of products. But the lack of response is a potential source of bias, and attention should be paid to any possible improvements of the present situation.

- **The discipline of the panellists.** Although some research users tend in some cases to ascribe a highly exaggerated accuracy to the data, you should not forget that they are based on the behaviour of a selected group of human beings. And you cannot be so naive to believe that all the panellists closely follows the instructions they have been given. You cannot believe that each and every panel member pushes the button every time he enters or leaves the room, when going to the rest-room, to the kitchen to drink some water, etc. If they did it, there would be sound reasons to question their capacity to represent the population, since that would not be the expected behaviour of normal human beings. My feeling is that, in this respect they basically tend to follow the rules, rules that they have adapted from the original instructions they have received, that adaptation of the rules being different from household to household. For instance, a panellist may feel obliged to sign off when he leaves the house but not when he is absent from the room for a short period. In short, the discipline is good enough to reasonably represent the individual behaviour in relation to TV watching, but there is plenty of room for improvement. Using a comparative evaluation, there is no doubt that the approximation to the watching behaviour is by far closer in the meter systems than in the yesterday-recall procedures.

As an overall evaluation, my opinion is that the Sofres AM operation is reasonably correctly operated and reflects quite adequately the actual levels of TV audience and its distribution by channels.

To conclude, let me tell you of a case when the meter system led to aberrant audience estimations. About a year ago, a national TV station announced it would broadcast a football match between the two important teams of Seville. The contest, a friendly game, took place to raise funds for the operation in the US of a 2-year-old boy sick with leukaemia. The TV station would contribute 18 million pesetas plus an additional and variable amount depending on the audience of the match, one peseta per viewer (average audience across the match length). The channel's offer was highly advertised by the different media during the days prior to the game.

The audience data obtained were extraordinarily high, above all expectations, and surpassed even that of high-interest matches of the national team. What had happened? The message of solidarity went straight to the hearts of many panellists.

They reasoned as follows. The audience is measured through a group of households I belong to, so I have in my hands the possibility of raising more money for the poor sick child. What shall I do? I must declare that I am watching the game on TV.

There were some extreme households where, according to the meter's data, the whole match was watched by the entire family, from the 3-year-old girl to the 85-year-old grandmother, plus three or four guests.

Everything indicated that the figures were exaggerating the actual audience level. The case shows up to what extent the panel members are conscious of the important role they play. And from the research side, this consciousness was neither looked for nor

wanted. This story reflects one negative aspect arising from the human side of the meter methodology.

Due to lack of time, I'll limit myself to mention the tests or activities will only go into some detail for some of them. I hope you will forgive me if I treat some points too briefly.