THE VALUE OF COINCIDENTAL SURVEYS TO MONITOR THE VALIDITY OF TV METER PANEL MEASUREMENTS

EMRO Conference 2005, Kwa Maritane, South Africa.

Carlos Lamas, AIMC

Introduction

The coincidental methodology emerged early on in the history of systems of audience measurement for the electronic media (radio and television). No one knows for sure who invented it or who used it first, but we do know that *George Gallup* measured radio audiences using face-to-face coincidental interviews at *Drake University* over the period 1921-1931, and that the first telephone coincidental study to publish its results was that by *John Karol* for the *Yankee Network* station in Boston in 1932.

However, there is no doubt that merit for popularising the technique goes to *C.E Hooper*, who launched the first continuous syndicated measurement operation in US radio in 1938, using coincidental telephone interviews. Criticising the recall technique used by CAB, which provided the radio ratings of the time, on the grounds of the assumed weaknesses of memory when confirming past listening (i.e. respondents forget, have memory lapses, or make potentially biased statements), he established a system that relied on three basic questions:

- *Are you listening to the radio?*
- *What programme?*
- *On what station?*

The *Hooperatings* set the pace for radio history up until the end of the 40s, when *Hooper* could no longer withstand the competition of the *Nielsen Radio Index* ratings. US radio is currently measured by *RADAR* which uses a diary-based methodology.

The coincidental technique does not rely on memory, as the listener is being asked to report on which is happening at the time. This immediacy would also seem to make listeners lest susceptible to any possible bias.

However, for a number of reasons the coincidental method is no longer viable as a continuous measurement system. These reasons include its high cost (it requires representative samples for each of the time intervals it aims to measure), its time limitations, as it cannot collect data from night-time or early-morning audiences, and because it cannot -mobile telephones notwithstanding- be used to gather data on people listening to the radio in their car.

In May 1963 CONTAM (Committee on National Television Audience Measurement) was set up, comprising representatives of the three national channels –ABC, CBS and NBC– the NAB (National Association of Broadcasters), in order to carry out a series of methodological studies on measurement systems. For its **study number 4**, with the help of SRI (Statistical Research Inc.), CONTAM designed a coincidental study with 4,000 respondents randomly selected from telephone directories in April 1969. Special care was taken to determine the precise number of rings before accepting a *no answer*, the construction of the questionnaire, the training and supervision of the interviewers, etc. A very great effort was made to determine the nature of the numbers that did not answer the call in the time interval assigned by means of call backs. This meant that there were numbers that were called up to 30 times for as long as three weeks after the initial call. The aim of this operation was to establish whether the number was a household, business, disconnected line, etc. and so be able to perform a formally correct evaluation of the results. Only in the case of 1.5% of the numbers originally chosen was it not possible to establish their nature. Numerous second calls were made to people who had initially refused to participate, offering them a financial incentive this time around, and numbers that were initially engaged were also redialled. This rigorous coincidental technique was

again used by CONTAM and SRI in their **studies 5 and 6,** and under the name Industry Sponsored Research (ISR) telephone coincidental it became the method against which any other system of measurement had compare itself in order to verify its validity. This "gold standard" feature of coincidental is consecrated in the *Principles of Nationwide Television Audience Measurement* published in 1990.

The GGTAM document and coincidental studies

In 1999, as a result of the efforts made by various international professional associations, an important document entitled "Towards Global Guidelines for Television Audience Measurement" was published. This was the outcome of work by a multinational, multidisciplinary team –the Audience Research Methods Group– led by the American Gabe Samuels, from ARF, and coordinated by the independent British consultant Peter Menneer. It offered a vision of the practices recommended for each of the different aspects of a television measurement operation, specifically using peoplemeters.

Section 20 of chapter VI (Data Collection) is devoted to coincidental studies. Firstly, it describes two basic methodologies for coincidental studies:

- Internal coincidental studies, the aim of which is to evaluate the degree of discipline of the members of the households on the panel regarding their obligation to identify themselves on the peoplemeter by pressing the relevant button on the remote control when they are watching television. This is checked by making telephone calls to the panellist households and asking how many television sets are turned on, the number of people watching each and the channel they are watching at that moment. These data are compared with the data from the peoplemeters to produce indicators of consistency.
- External coincidental studies, the aim of which is to evaluate the reliability of the peoplemeter measurement as a whole by comparing its results with those obtained from a coincidental telephone study of a sample of households totally independent from the sample in the peoplemeter panel.

At the same time, GGTAM recommends performing a coincidental study at least once a year, but with the vagueness typical of some parts of the document, does not specify whether it should be one of each type (i.e. one internal and the other external) or if one of the types is sufficient.

Although GGTAM goes into the subject of coincidental studies and recommends their use (it should be noted that the document on *Minimum Standards For Media Rating Research* issued by the American *Media Rating Council* does not even mention them), does not give its unqualified blessing to external coincidental studies as the "gold standard".

Internal coincidental studies in Spain

Since 1988 a telephone coincidental study has been carried out each year on households in the panel. The main characteristics of these studies are:

- The aim has always been to reach all the households on the panel, although for one reason or another not all of them can be taken into account in the analysis.
- Calls are made during a period of two consecutive weeks, including weekends.

- To increase the probability that households have one or more television sets turned on at the time of the call, interviews are concentrated at daily audience peaks (13.00-16.00 and 19.00-22.00).
- The calls are distributed evenly over the times and days of the study. It might seem better to distribute samples over time in proportion to the audience at each time, but for practical reasons (to maintain an even workload) the choice of constant size samples in each time unit is preferred.
- For a time two focuses were used in parallel: in half of the sample the interviewer identified himself or herself as an employee of the meter panel operator (TNS) and said that he or she was running a check on the operation of the meter; in the other half of the sample the interviewer simply said that a given research institute (not TNS, of course) was conducting a survey on the use of the media (and questions on the press and radio were also included). The results of the first subset of the sample were systematically more favourable, but the differences were of limited importance. As of the year 2000 study, it was decided that only the former focus would be used so as to increase the efficiency and quality of data collection, and the slight bias -already known- inherent to this approach was accepted.
- A CATI system is used and the comparison between the information from the interviews and the peoplemeter logs is highly computerised.
- The design of the study, and its implementation, are subject to the supervision of AIMC[1] (the external auditor of the TV measurement operation).
- The questionnaire used covers:
    - The people who are in the home at the time of the call (members of the household and guests).
    - Televisions that are turned on at the time the telephone began to ring, and the channels and programmes being watched on each.
    - The people who are in the rooms where there is a television switched on.
    - The people who were watching the television at the time the telephone rang. In the case of guests, their age and sex.

The results are usually presented in tabular form, where the information from the telephone study is compared with the peoplemeter data. In the case of the 2004 study, this table was:

[1] Asociación para la Investigación de los Medios de Comunicación

Telephone survey

		Claiming to view TV	Claiming not to view TV	Total
Peoplemeter	Button pressed	2005 24.7%	500 6.1%	2505 30.8%
	Button not pressed	373 4.6%	5235 64.6%	5608 69.2%
	Total	2378 29.3%	5735 70.7%	8113 100%

Two indicators are habitually used to summarise the results of the coincidental comparison. These are:

- *Full coincidence index* (total percentage of individuals for whom survey data was consistent with the peoplemeter identification: 89.3% (24.7+64.6).
- *Peoplemeters/coincidental ratio,* calculated as the ratio between the peoplemeter audience and the coincidental audience.

$$\text{Peoplemeters/coincidental ratio} = \frac{2505}{2378} * 100 = 105.3\%$$

The coincidence index calculated in this way is not very robust as it is strongly influenced by the selection of time slots. In other words, if the coincidental study were to be performed between 10 and 11 a.m., in most homes the television would be turned off, there would be very little audience registered and the meters would not report any audience either. This would lead to a very high level of coincidence, without having really judged the level of discipline of the panellists when pressing the buttons on their meter. It therefore seems reasonable to look for alternatives that are able to overcome this weakness.

If we call the index described above "Coincidence Index I", we could define two alternative indicators in the following way:

Index II: Defined as the percentage that the television viewers according to both systems represent of the total number of television viewers according to the coincidental survey.

$$\text{II} = \frac{2005}{2005 + 373} = \frac{2005}{2378} = 84.3\%$$

Index III: Defined as the percentage that the television viewers according to both systems represent of the total number of television viewers according to the either of the two measurements.

$$\text{III} = \frac{2005}{2005 + 373 + 500} * 100 = \frac{2005}{2878} * 100 = 69.7\%$$

The time course of these three indices over the last 12 studies was as follows:

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Index I	87.8	89.6	88.4	90.3	90.1	89.7	90.3	90.3	89.3	90.9	89.2	89.2
Index II	81.8	83.4	82.2	83.6	82.9	83.6	85.3	87.1	83.3	84.8	83.6	84.3
Index III	68.5	72.9	70.8	72.9	71.6	72.2	73.6	71.4	69.9	74.2	69.4	69.7

And the time course of the peoplemeter/coincidental ratio,

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Peop/coin. ratio	101.3	97.8	98.3	98.3	98.7	99.3	101.0	109.3	102.5	99.1	104.0	105.3

While the series of coincidences, in particular that shown by Index I shows significant stability, the list of audiences offered by the peoplemeter/coincidental ratio is somewhat irregular and even changes sign.

It is worth noting the comparison from the point of view of households. It is observed that the full coincidence of households where at least one individual was watching the television represented only 40% of all homes taking part in the trial. And almost another 40% were households where the coincidence was less noteworthy because nobody was watching television at the time of the call.

	2002		2003		2004	
	Cases	%	Cases	%	Cases	%
Total homes monitored	2763	100.0	2743	100.0	2776	100.0
Full coincidence of individuals	2239	81.0	2115	77.1	2170	78.1
Are watching television	1219	44.1	1141	41.6	1111	40.0
Are not watching television	1020	36.9	974	35.5	1059	38.1
Numeric, but not full, coincidence	87	3.2	99	3.6	85	3.1
Non numeric coincidence	437	15.8	529	19.3	521	18.8
More viewers on meter	214	7.7	303	11.1	319	11.5
Less viewers on meter	223	8.1	226	8.2	202	7.3

What is very significant is the analysis relating to guests (guests were excluded from the analysis reported so far because although measured, their viewing is not included in the standard information. The table of coincidence in the 2004 study shows the following situation:

Telephone survey

		Claiming to view TV	Claiming not to view TV	Total
Peoplemeter	Button pressed	43 16.7%	25 9.7%	68 26.5%
	Button not pressed	96 37.4%	93 36.2%	189 73.5%
	Total	139 54.1%	118 45.9%	257 100%

It may be observed that Coincidence Index I only reaches 52.9% and that the audience measured by the meter is only 48.9% of the audience reported by the telephone survey. By all measures the real audience is underrepresented by the peoplemeter measurement. Panellists are probably not entirely conscious of their obligation to identify guests on the meter.

One of the by products that is also obtained from the results of the coincidental analysis is a measure of the congruence between the channel reported in the interview for each television set and the channel recorded by the meter. Although they represent a small number of cases (around 3% of all cases in the 2004 study) the data could be significant if they point to errors in the measurement operation. However, we should point out that historical experience after many years of verifications clearly highlights that the errors tend to lie in the channel stated during the telephone interview (confusion in the declaration or incorrect noting down) and only in a few instances are can they be put down to the peoplemeter operation.

In Spain, panellists are asked to identify themselves on the meter when they consider that they are “watching the television”, in contrast with other countries where the criterion used is that of being “in a room where there is a television set turned on”. Additionally, the internal coincidental study provides us with an estimate of the quantitative impact of using one criterion or another. The table shown below comes from the 2004 study.

		In room with TV turned on according to survey			
		Yes.		No	
		Claiming to view TV	Claiming not to view TV		Total
Peoplemeter	Button pressed	2005 24.7%	66 0.8%	434 5.3%	2505 30.9%
	Button not pressed	373 4.6%	120 1.5%	5115 63.0%	5608 69.1%
	Total	2378 29.3%	186 2.3%	5549 68.3%	8113 100%

Of the 2564 members of households who were in rooms where there was a television set turned on, 93% described themselves as television viewers. Of the 7% who did not, a third had pressed the relevant button on the peoplemeter.

External coincidental studies in Spain

One Sunday in January of 1991, the national newspaper ABC published the list of panellist households used for the television audience measurement, at that time operated by the company ECOTEL. Although the published list was not entirely up-to-date, the identity of 75% of the households on the panel at the time was unveiled. ECOTEL held a meeting with representatives of all the television channels to discuss a plan of action, based on the premise that the panel would have to be completely replaced. The main dilemma faced was deciding whether it would be better for the market to do without the audience data while the new panel was being built, or to continue to use and accept the audience data as they had done so far. The group's common sense led it to decide to continue with the measurements. The television channels were to support the validity of the information from the disclosed panel, and they undertook not to contact the households on the published list. However, they required ECOTEL to undertake three external coincidental studies in 1991 with which to contrast the data from the panel and confirm the reliability of the data from them.

The first two coincidental studies were carried out using what we might call the classic methodology. That is to say, they took a representative sample of homes in the target population distributed by hours and days, and from homes that answered the telephone call they collected data on the number of members of the household that were watching the television, the characteristics of the household and its members. Homes that did not answer in their allotted time interval were called back again to determine their type (uninhabited dwelling, normal home, business premises, etc.) From all this the total television audience in the telephone study and their share of each of the channels was estimated. After the two first external coincidentals I proposed - I was then working for ECOTEL as Technical Director- a change in the methodology for two reasons:

- What the market, and the television channels in particular, was interested in was checking possible biases in favour of a given channel and not verifying whether the total television viewing data were correct. That is to say, they accepted the idea of only monitoring each channel's viewing share.
- The additional objective of measuring total television viewing resulted in marked inefficiencies in the study. These included interviewing homes that did not add anything to the calculation of viewing share as they were not watching television at the time of the call, a high number of calls back only to establish whether the number really corresponded to an inhabited main residence, etc.

The alternative methodology put forward by the Users' Committee consisted of taking a sample of homes in each time interval that was representative of the homes where there was at least one person watching the television (not necessarily representative of the general population). It relied on the plausible hypothesis that the homes that did not answer the telephone were not watching television at the time (because they were not at home) and, therefore, are not part of the target sample. Households that answered but which were not watching television did not enter the sample either. All the homes that were included in the sample yield information for the calculation of viewing share. Given that you need to combine results from different half-hour periods with different total television audience levels, it is inevitable that the total television viewing in each time interval will be introduced into the calculation of the expansion factors. This total television viewing for each of the time intervals is taken from the information provided by the peoplemeter panel (data that, in general, is not at all questioned by the market). The audience characteristics in each time interval are also collected from the panel (social class, ages and sex, regional language, size of household, etc.) so they can be included into the rim-weighting program. The first study using this new approach was conducted by IOPE in June 1991, and since then all subsequent coincidental studies have followed suit.

The questionnaire basically includes, for each home interviewed, the number of television sets switched on at the time of the call, the channel and programme being watched on each of these television sets and the people in front of each set, with their characteristics: sex, age, socio-economic status and others.

The table below gives a list of the coincidental studies that have been supervised by AIMC up to 2004. Most of these studies were commissioned and financed by the measurement operator (first ECOTEL and later TNS), but the list also includes three studies undertaken by two important national channels: Antenna 3 and TVE.

	Commissioned by	Institute responsible	Sample size (households)	Sample size (individuals)
1991, Mar	**Ecotel**	**Demoscopia**	4,957	
1991, Apr	**Ecotel**	**IOPE**	4,033	
1991, Jun	**Ecotel**	**IOPE**	6,623	11,010
1993	**TNS AM**	**IOPE**	8,609	16,312
1994	**TNS AM**	**IOPE**	8,425	15,533
1995	**TNS AM**	**IOPE**	8,400	15,502
1996	**TNS AM**	**IOPE**	8,398	14,823
1997	**TNS AM**	**IOPE**	9,444	16,768
1998	**TNS AM**	**Metra Seis**	9,450	17,164
1999	**TNS AM**	**Metra Seis**	9,452	16,512
2000	**TNS AM**	**Metra Seis**	9,465	16,506
2001	**TNS AM**	**Metra Seis**	9,450	16,083
2002, May	**TNS AM**	**Metra Seis**	9,458	15,987
2002, Nov	**Antena 3**	**EMER-GFK**	7,555	13,314
2003, Jan	**TNS AM**	**Metra Seis**	9,432	16,617
2003, Feb-Mar	**Antena 3 and TVE**	**EMER-GFK**	7,546	12,921
2003, May	**TNS AM**	**Metra Seis**	7,515	11,891
2004	**Antena 3**	**EMER-GFK**	9,406	15,547

The samples are established based on the number of households (always those watching television) and the figure for the sample of individuals appearing in the last column is the number of people watching television at the time of the call.

In the institutional studies (i.e. those commissioned by the measurement operator), the Users' Committee is responsible for the selection of time slots in which to perform the study. Historically, two intervals corresponding to periods of maximum audience have been used: 14:00-16:30 and 20:00-22:30. It is standard practice not to notify the market in advance of the dates on which the study will take place (this is agreed by TNS and AIMC). Recently it has also been decided that TNS is not to be informed of the dates so as to ensure there are no activities which might alter the normal course of the study and the subsequent comparative analysis. Telephone interviews are normally performed over the course of a full week.

The distribution of the sample over half-hour intervals and days is strictly uniform. However, for both statistical and logistic reasons it is advisable to distribute the sample with a probability proportional to the total television audience. This improves the homogeneity of the expansion factors, which has a positive effect on the sampling efficiency as well as spreading the workload more evenly over the various time slots (with a uniform distribution, at times of low relative audience, a larger number of interviewers is needed to achieve the sampling target -many calls do not contribute to the usable sample- than at times of higher audience).

The results of the coincidental studies carried out over for the period 1991-2004 are presented below. The table shows the Total Spain results for all the time slots covered, in terms of differences with the audience shares given by the peoplemeters. The differences are calculated as *peoplemeter share - coincidental share.* Therefore, a positive sign difference has to be understood as implying that meter operation gives a higher share than the coincidental study.

	TVE 1	La 2	Tele 5	Antena 3	Canal Plus	Regional	Others Channels
1991, Mar	-0.5	1.2	-2.0	-----	-----	0.9	0.4
1991, Apr	0.5	2.7	-2.5	-1.4	0.2	0.8	-0.3
1991, Jun	1.3	2.1	-2.5	0.0	-0.1	-0.1	-0.7
1993	0.3	0.0	-1.1	0.2	-0.1	0.8	-0.1
1994	-0.5	1.4	-1.5	-0.2	0.0	1.1	-0.3
1995	-2.1	1.2	1.2	-0.2	-0.8	0.5	0.2
1996	-1.4	0.5	1.7	0.2	-1.4	0.2	0.2
1997	-1.5	0.1	0.2	1.1	-0.5	0.3	0.3
1998	-2.0	0.9	0.2	0.9	-0.6	0.6	0.0
1999	-1.0	0.3	0.7	0.1	-0.4	0.5	-0.1
2000	-1.8	0.6	1.0	-0.8	0.1	0.6	0.2
2001	-1.1	0.4	0.8	-0.5	0.0	1.1	-0.7
2002, May	-2.6	0.2	1.0	0.5	-0.2	1.5	-0.5
2002, Nov	-2.8	0.4	3.0	-0.1	0.2	-0.3	-0.4
2003, Jan	-2.5	0.9	1.5	0.1	0.2	0.2	-0.4
2003, Feb-Mar	-1.3	0.0	1.0	0.9	0.2	-1.1	0.0
2003, May	-3.3	1.4	1.0	1.0	-0.2	0.0	0.0
2004	-4.5	0.5	0.6	2.0	-0.6	0.5	1.5

Difference=Share (Peoplemeters) - Share (Coincidental)

One fact that really stands out in the series is the degree of disagreement in the comparative analysis of 2004, breaking with certain stability in institutional coincidental studies, with a difference of 4.5 points for TVE1. While it is true that for TVE1 the sign of the differences has been systematically negative since 1994, the magnitude in previous years was always smaller, never exceeding a difference of 3 points. It did rise to three percentage points in the last coincidental study commissioned by Antena3 in May 2003. No doubt, the systematic negative difference for TV1 is, from a historical perspective, the most worrying factor. This concern is heightened further by the observation of the size of the differences for TVE1 in the last two studies.

In the case of the rest of the channels the situation is much less worrying. The differences in La2, which are always positive, are much smaller. And in the case of other stations the differences almost never reach levels sufficient to be a cause for concern. Moreover, the alternation between positive and negative values suggests that a systematic bias by the panel cannot be inferred.

Every year the differences are evaluated both statistically terms and on the basis of the margins of error of each of the two studies compared. The tables of data published showing the results of this comparison are of the type shown below (the tables show the results of the last two studies).

COMPARISON OF CHANNELS' VIEWING SHARES. MAY 2003

	EMER-GFK		SOFRES AM		DIFFERENCE			
	SHARE	ST. ERROR	SHARE	ST. ERROR	SHARE	ST.ERROR	K	
TVE1	27.7	0.6	24.4	0.4	-3.3	0.8	-4.4	**
La 2	4.6	0.3	6.0	0.2	1.4	0.4	3.8	*
Tele 5	19.6	0.6	20.6	0.4	1.0	0.7	1.5	
Antena 3 (T)	18.5	0.6	19.5	0.4	1.0	0.7	1.5	
C.Plus (T)	4.4	0.3	4.2	0.2	-0.2	0.3	-0.7	
Regional	18.9	0.6	18.9	0.4	0.0	0.7	0.1	
CSD (sin C+)	1.3	0.2	1.3	0.1	0.0	0.2	-0.1	
VD (sin A3)	0.9	0.1	1.0	0.1	0.1	0.2	0.7	
Other	4.2	0.3	4.1	0.2	-0.1	0.3	-0.3	

SAMPLE SIZE INDIVIDUALS	
EMER-GFK	12268
SOFRES AM	9424

COMPARISON OF CHANNELS' VIEWING SHARES. APRIL 2004

	METRA SEIS		TNS AM		DIFFERENCE			
	SHARE	ST. ERROR	SHARE	ST. ERROR	SHARE	ST.ERROR	K	
TVE1	28.5	0.6	24.0	0.4	-4.5	0.7	-6.4	**
La 2	5.9	0.3	6.4	0.2	0.5	0.4	1.3	
Tele 5	20.8	0.5	21.4	0.3	0.6	0.6	0.9	
Antena 3	17.7	0.5	19.7	0.4	2.0	0.6	3.2	*
C.Plus (T)	3.9	0.3	3.3	0.2	-0.6	0.3	-2.0	*
Regional	17.8	0.5	18.3	0.3	0.5	0.6	0.8	
THEMATIC	3.1	0.2	4.8	0.2	1.7	0.3	5.8	**
Others	2.2	0.2	2.0	0.1	-0.2	0.2	-0.9	

SAMPLE SIZE INDIVIDUALS	
METRA SEIS	15547
TNS AM	9244

The standard error of the difference is calculated as the square root of the sum of the squares of the errors of the two studies. The ratio between each of the differences and its corresponding standard error is denoted by *K.* Differences for which the value of *K* is greater than 2 (outside the confidence interval constructed with a 95% probability and centred on the null hypothesis, i.e. difference = 0) are usually considered statistically significant. The differences with a value of *K* greater than 2 are marked with an asterisk.

However, we are aware that we are not looking at studies that are impeccable in terms of their sampling process. Nor can the differences entirely be put down to the fact that the samples are different, as the methodology used to obtain the information is different in each case and this always brings with it the possibility of additional differences caused by factors other than the sampling. For all these reasons, several years ago, the Users' Committee accepted a proposal I made to consider differences to be "really serious" when the value of *K* exceeded 4. These are marked in the table with two asterisks.

Neither of the methods compared here is perfect. In the meter panel we have a serious problem of response rate (around 5%) and the discipline of the panellists - who have to press buttons on a remote control- leaves a lot to be desired.

In the coincidental telephone study, we have historically had various problems arising from the lack of coverage of the sampling framework used (telephone directory): unlisted numbers, telephones on cable networks and -the most important problem of all, and one which has not yet been tackled- the zero probability of contacting that 20% of the population that has no fixed telephone in their home (i.e. households that rely exclusively on mobile phones). In addition, the rate of collaboration (calculated as the proportion of homes that collaborated compared to the total number of homes contacted and watching the TV) is only 58% (as obtained in the 2004 study). And it is undoubtedly true that the channel viewing declarations are not totally error free. Moreover, the same variation in the differences from year to year (when the panel remains partially constant) raises a question mark over the precision of coincidental studies. Altogether this suggests that, at least in Spain, we cannot consider coincidental studies to be the *gold standard.* At most we could talk of a *silver standard*☺.

However, we do not think that the differences regarding TVE1 can be explained by the defects of the sampling framework due to its not including households that rely exclusively on mobile phones. We have tabulated data from the TNS AM in two forms by selecting only those homes with a fixed telephone. The table below shows a comparison (using the May 2004 totals) of the regular panel data and those obtained from panellists with a fixed telephone, retaining the regular expansion factors, and lastly, expanding the homes with a fixed telephone to the total universe of households by reweighting the data.

EFFECT OF FIXED TELEPHONY ON CHANNELS' VIEWING SHARES (May 2004)

Slot	Channel	Total	Panel with fixed tel.	Panel with fixed tel.*
Total Coincidental	TVE1	23.8	24.6	22.9
	La2	6.4	6.4	6.5
	T5 (A)	21.1	20.8	22.7
	A3 (A)	20.1	19.3	19.8
	C+	2.9	2.9	2.0
	REG	18.3	18.9	17.5
	THEMATIC	5.1	4.9	5.7
	TV LOCAL	2.2	2.1	2.8
	OTHER	0.2	0.1	0.1
Total Day	TVE1	22.8	23.3	24.1
	La2	6.6	6.5	6.3
	T5 (A)	22.8	22.4	21.0
	A3 (A)	20.0	19.7	19.7
	C+	2.0	2.0	2.9
	REG	17.3	17.8	18.7
	THEMATIC	5.5	5.5	5.1
	TV LOCAL	2.9	2.8	2.1
	OTHER	0.2	0.1	0.1

* Expanded to the total population

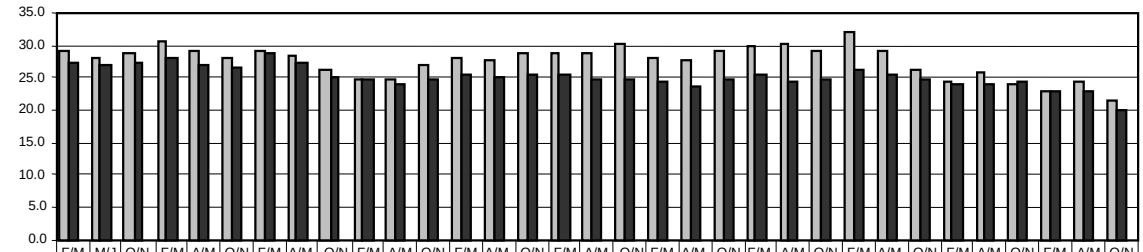
We see that in the time slot used in the coincidental study the expansion of the fixed telephone households in the panel to the total number of households (simulation of the process used in the coincidental study) gives a value for TVE1 that is yet smaller

than the regular data from the meters, undermining the possible hypothesis explaining the differences observed for TVE1 by suggesting that “mobile only” individuals (not represented in the telephone study) are more likely to watch this channel.

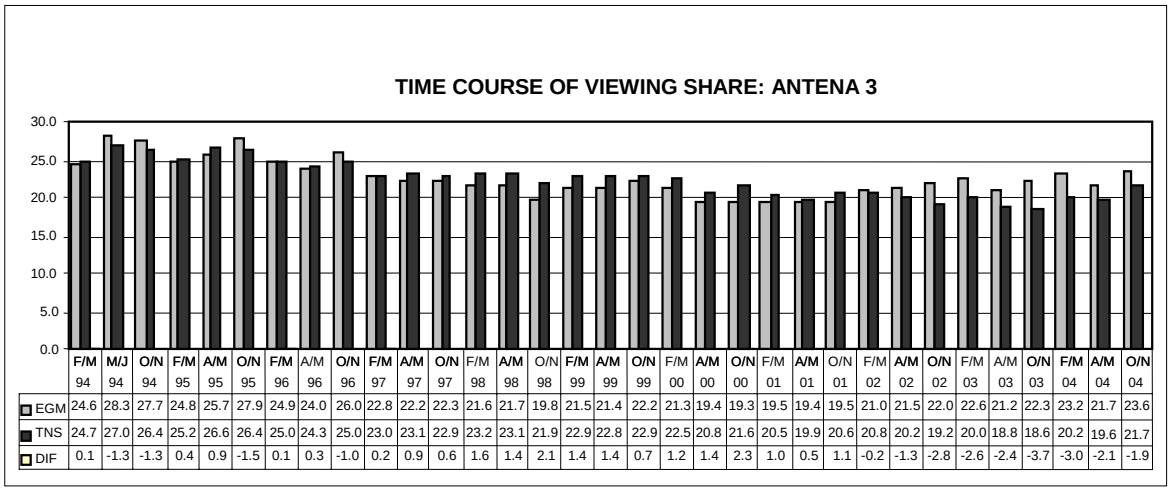
Also, assuming the absolute methodological superiority of the coincidental study can lead us onto very uncertain ground. It may happen that a marked difference in the comparison points to an objective weakness on the part of the panel. For example, a few years ago the comparison suggested that the panel was underestimating the audience of Canal+ and, at the same time, it was confirmed trough other controls that Canal+ subscribers were underrepresented on the panel. The corrective action to take, i.e. increasing the number of subscriber households, was obvious in this case. But that was somewhat exceptional. In general it is not possible to detect a fault with the panel (of any type, whether sampling, technological or operational) which can be blamed from the differences detected. So, what can be done? Of course, there are ways of approximating panel audiences to the profile indicated in the coincidental study (replacing households with a low level of TVE1 viewing, if TVE1 is the channel the panel underestimates, for others that watch more TVE1), but it is difficult for a corrective measure of this type (or any other that has a similar appearance of an artificial manipulation) would be agreed to by the other channels. We can find ourselves in a cul-de-sac where the only way out is to repeat the coincidental study until the figures it gives are more consistent with those of the panel.

One way of making progress in the analysis is to compare the panel’s audience results with other alternatives, where they exist. Fortunately, in Spain we have the *Estudio General de Medios (EGM)*, which uses personal interviews with CAPI to measure audience figures across various media, including television. It uses the “day before recall” technique with a sample of 43,000 interviews a year. AIMC also compares both sources regularly, as shown in the graphs below.

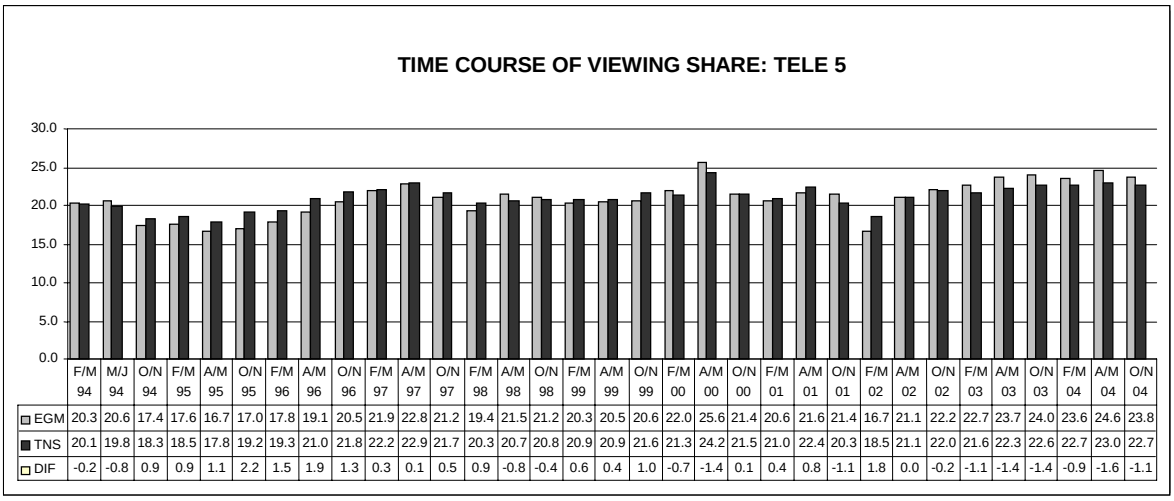
TIME COURSE OF VIEWING SHARE: TVE1



	F/M 94	M/J 94	O/N 94	F/M 95	A/M 95	O/N 95	F/M 96	A/M 96	O/N 96	F/M 97	A/M 97	O/N 97	F/M 98	A/M 98	O/N 98	F/M 99	A/M 99	O/N 99	F/M 00	A/M 00	O/N 00	F/M 01	A/M 01	O/N 01	F/M 02	A/M 02	O/N 02	F/M 03	A/M 03	O/N 03	F/M 04	A/M 04	O/N 04
EGM	29.1	27.9	28.7	30.6	29.3	28.1	29.0	28.2	26.4	24.7	24.7	27.0	28.0	27.6	28.8	28.7	28.8	30.3	28.1	27.6	29.2	29.8	30.3	29.3	31.9	29.3	26.2	24.2	25.9	23.8	23.0	24.4	21.4
TNS	27.4	27.0	27.1	28.1	27.0	26.7	28.9	27.4	25.1	24.8	23.9	24.9	25.3	25.1	25.5	25.3	24.7	24.9	24.4	23.7	24.7	25.5	24.3	24.7	26.4	25.5	24.8	23.8	24.1	24.4	23.0	23.0	20.0
DIF	-1.7	-0.9	-1.6	-2.5	-2.3	-1.4	-0.1	-0.8	-1.3	0.1	-0.8	-2.1	-2.7	-2.5	-3.3	-3.4	-4.1	-5.4	-3.7	-3.9	-4.5	-4.3	-6.0	-4.6	-5.5	-3.8	-1.4	-0.4	-1.8	0.6	0.0	-1.4	-1.4



	F/M 94	M/J 94	O/N 94	F/M 95	A/M 95	O/N 95	F/M 96	A/M 96	O/N 96	F/M 97	A/M 97	O/N 97	F/M 98	A/M 98	O/N 98	F/M 99	A/M 99	O/N 99	F/M 00	A/M 00	O/N 00	F/M 01	A/M 01	O/N 01	F/M 02	A/M 02	O/N 02	F/M 03	A/M 03	O/N 03	F/M 04	A/M 04	O/N 04
EGM	24.6	28.3	27.7	24.8	25.7	27.9	24.9	24.0	26.0	22.8	22.2	22.3	21.6	21.7	19.8	21.5	21.4	22.2	21.3	19.4	19.3	19.5	19.4	19.5	21.0	21.5	22.0	22.6	21.2	22.3	23.2	21.7	23.6
TNS	24.7	27.0	26.4	25.2	26.6	26.4	25.0	24.3	25.0	23.0	23.1	22.9	23.2	23.1	21.9	22.9	22.8	22.9	22.5	20.8	21.6	20.5	19.9	20.6	20.8	20.2	19.2	20.0	18.8	18.6	20.2	19.6	21.7
DIF	0.1	-1.3	-1.3	0.4	0.9	-1.5	0.1	0.3	-1.0	0.2	0.9	0.6	1.6	1.4	2.1	1.4	1.4	0.7	1.2	1.4	2.3	1.0	0.5	1.1	-0.2	-1.3	-2.8	-2.6	-2.4	-3.7	-3.0	-2.1	-1.9



	F/M 94	M/J 94	O/N 94	F/M 95	A/M 95	O/N 95	F/M 96	A/M 96	O/N 96	F/M 97	A/M 97	O/N 97	F/M 98	A/M 98	O/N 98	F/M 99	A/M 99	O/N 99	F/M 00	A/M 00	O/N 00	F/M 01	A/M 01	O/N 01	F/M 02	A/M 02	O/N 02	F/M 03	A/M 03	O/N 03	F/M 04	A/M 04	O/N 04
EGM	20.3	20.6	17.4	17.6	16.7	17.0	17.8	19.1	20.5	21.9	22.8	21.2	19.4	21.5	21.2	20.3	20.5	20.6	22.0	25.6	21.4	20.6	21.6	21.4	16.7	21.1	22.2	22.7	23.7	24.0	23.6	24.6	23.8
TNS	20.1	19.8	18.3	18.5	17.8	19.2	19.3	21.0	21.8	22.2	22.9	21.7	20.3	20.7	20.8	20.9	20.9	21.6	21.3	24.2	21.5	21.0	22.4	20.3	18.5	21.1	22.0	21.6	22.3	22.6	22.7	23.0	22.7
DIF	-0.2	-0.8	0.9	0.9	1.1	2.2	1.5	1.9	1.3	0.3	0.1	0.5	0.9	-0.8	-0.4	0.6	0.4	1.0	-0.7	-1.4	0.1	0.4	0.8	-1.1	1.8	0.0	-0.2	-1.1	-1.4	-1.4	-0.9	-1.6	-1.1

As may be observed, EGM also systematically reports a larger audience for TVE1 than does TNS. However, the differences have been less marked in recent periods, as is shown by the comparison of the last coincidental survey with the peoplemeter measurement. This relationship is easier to understand because usually the recall method favours the best-known channels. In other words, EGM does not support the thesis that the panel underestimates the audience of TVE1.

The discrepancies observed in the 2004 coincidental study gave rise to considerable concern in the Spanish market. A number of hypotheses were suggested to explain the differences, but with little success. After much debate, the only viable option was to run another coincidental study after undertaking a thorough revision of the methodology used to date. The result of this revision was the introduction in the new study of a number of minor changes in the questionnaire. These made the sequence of questions a bit more logical and made the questionnaire somewhat shorter. Particular emphasis was also placed on ensuring that respondents were able to identify the channel they were watching by installing television sets in the interviewers´ room so that they could follow what was on the main channels.

The new study also expanded the range of time intervals to be studied, covering eight hours instead of the traditional five. The sample size was increased to 15,777 households and 22,850 individuals, the rim-weighting conditions were modified slightly and the field work was entrusted to two different institutes (Metra Seis and IMOP) to assess the impact of a possible “institute effect”.

The field work took place in the week from 11 to 17 January 2005. The results were as shown below.

COMPARISON OF CHANNELS’ VIEWING SHARES. JANUARY 2004

	METRA SEIS+ IMOP		TNS AM		DIFFERENCE			
	SHARE	ST. ERROR	SHARE	ST. ERROR	SHARE	ST.ERROR	K	
TVE1	22.0	0.4	20.3	0.4	-1.7	0.6	-2.7	*
La 2	5.2	0.2	6.0	0.1	0.8	0.3	2.8	*
Tele 5	21.2	0.4	22.0	0.4	0.8	0.6	1.3	
Antena 3	21.6	0.4	20.6	0.3	-1.0	0.6	-1.9	
C.Plus (T)	2.8	0.2	2.4	0.1	-0.4	0.2	-1.7	
Regional	18.1	0.4	19.2	0.5	1.1	0.6	1.9	
THEMATIC	6.4	0.3	6.5	0.4	0.1	0.5	0.1	
Other	2.7	0.2	2.9	0.1	0.2	0.2	1.0	

SAMPLE SIZE	
INDIVIDUALS	
METRA SEIS+ IMOP	22850
TNS AM	8852

We were lucky. The size of the differences was significantly reduced (both for the measurement period as a whole and for the traditional time slots) and, until the results of the next coincidental study are in, peace of mind has returned to the market. Nevertheless, the sign of the differences remains the same, and there is no plausible hypothesis to explain it.

The differences between the separate estimates for each of the institutes were not significant.

	IMOP		METRA SEIS		DIFFERENCE		
	SHARE	ST. ERROR	SHARE	ST. ERROR	SHARE	ST.ERROR	K
TVE1	21.7	0.7	22.5	0.7	0.8	0.9	0.9
La 2	5.3	0.4	5.1	0.4	-0.2	0.5	-0.4
Tele 5	21.4	0.6	21.0	0.7	-0.4	0.9	-0.4
Antena 3	21.5	0.6	21.5	0.7	0.0	0.9	0.0
C.Plus (T)	2.6	0.2	3.0	0.3	0.5	0.4	1.3
Regional	17.9	0.6	18.1	0.6	0.2	0.9	0.3
THEMATIC	6.9	0.4	6.0	0.4	-0.9	0.6	-1.6
Other	2,7	0.3	2.7	0.3	0.0	0.4	-0.1

SAMPLE SIZE INDIVIDUALS	
IMOP	11649
METRA SEIS	11201

International practices

It is not easy to know what is being done in this area in other countries. It is worth highlighting that the valuable compilation by *Toby Syfret* in his book "*Television Peoplemeters in Europe*", which describes a wide variety of characteristics of European measurement operations, does not devote a single comment to the issue.

Even without a detailed knowledge of what is happening in other countries, I think it is safe to assume that the practice of internal coincidental studies is fairly widespread, and that they almost always return reasonably high values of coincidence (between 86 and 95%) and do not cause much concern in the various markets involved. The essentials of the methodologies used are the same, although there are some variations in the procedures: total panel vs a sample of the panel, control of one person per home vs all members of panel (with or without guests), calls throughout the day or during specific time intervals, etc.

The case is different as far as external coincidental studies are concerned. In many countries they are simply not done, and in others the policy of disclosure of the results of those that are carried out is very restrictive. In short, I do not know of any country other than Spain that has an open policy of annual external coincidental studies. And it is highly significant that the very complete TAM[2] Glossary, published by the AGB Group on its web site (www.agbgroup.com) says that "*external coincidental surveys are rarely used in TAM research".*

According to the information on AGF's website (www.agf.com), external coincidental studies are carried out in Germany every two years. They use a sample of over 8000 homes and a CATI system to perform telephone interviews. The time slot for studies is between 18:25 and 20:40. They also run an internal coincidental study every two years, using a sub-sample of 1300 households on the panel in a time slot from 19:00-21:15. The 2000 study yielded a coincidence index of 90.7.

Someone once told me that in the United States they had stopped doing external coincidental studies as they caused confusion in the market and they did not know

[2] TAM: Televisión Audience Measurement

what to do with the results. Nor are they undertaken in France, the Netherlands or the United Kingdom –at least on a regular basis and with published results.

According to the information available to me, external coincidental studies as we know them in Spain have never been carried out in Italy. However, under the name of *"external coincidental check"* they perform an unusual exercise to demonstrate that the panel is not affected by bias resulting from the high level of refusal to participate. Households are recruited for the panel from those interviewed in its *"establishment survey"*. Obviously, not all the households contacted in this study agree to participate. The exercise therefore consists of contacting the households that declined to be interviewed in the "*establishment survey*" to verify that their television viewing habits are basically the same as those of the households on the panel. In 2002 a total of 3,006 households were interviewed by telephone during the 20:45 to 21:30 interval to ask them what channel they were watching at the time of the call. The audience share figures for the various channels obtained in this study were compared with those from the panel, confirming them to be basically the same.

Finally, in Argentina various controlled external coincidental studies have been conducted by the CCMA[3]. The results have been published and generated considerable controversy. Argentina is probably the country in which the external coincidental technique has had greatest importance as a control for the peoplemeter operation and where the discussions, analyses and hypotheses regarding the reasons for the differences have been richest and most open.

Conclusions

Internal coincidental studies are standard practice in most countries. They are carried out by the measurement operator and the indices of coincidence are, with some exceptions, fairly similar. Their utility is clear, and their results are not normally controversial.

The use of external coincidental studies as a control for the operation of peoplemeters is much less frequent. Although a high degree of congruence between these studies and the results of peoplemeters studies is to be expected, it should not be forgotten that there is a general principle that says that when the methods change, the results they produce also change. The different methodologies are different instruments –all of them imperfect—which aim to approach the "truth" from different angles. Moreover, it seems to me difficult to maintain that, in the current situation, the coincidental technique can be considered the ultimate method against which to benchmark other methods. Nevertheless, it would seem highly advisable and "good hygiene" to undertake external coincidental studies at regular intervals to detect possible biases or malfunctioning in the panel as there is no other better tool with which to evaluate the overall reliability of peoplemeters measurements.

[3] Cámara de Control y Medición de Audiencia, a similar organisation to Spain's AIMC.

Bibliographic References

- AIMC. *El control de audimetría en España (1993-2004).*
- Audience Research Methods Group. *Towards Global Guidelines for Television Audience Measurement*. EBU, Geneva, 1999.
- Beville Jr, Hugh Malcom. *Audience ratings*. Lawrence Erlbaum Associates, Publishers. New Jersey, 1985.
- Danaher, Peter and Beed, Terence. *A Coincidental Survey of People Meter Panelists: Comparing What People Say with What They Do*. Journal of Advertising Research. January/February, 1993.
- Lamas, Carlos. *El valor de los estudios coincidentales como contraste de la medición audimétrica*. 21º Seminario AEDEMO de Televisión. Valencia, February 2005.
- Media Rating Council, Inc. *Minimum Standards For Media Rating Research*. New York, October 1997.
- Scaglia, Dominique and Braun, Jacques. *TV audience measurement validation: present and future*. ESOMAR Seminar "The Quality of Media Information". Paris, April 1990.
- Smit, Piet. *SAARF TAMS coincidental study and independent audit results*. EMRO Conference 2002, Crete.
- SRI/CONTAM. *Principles of Nationwide Television Audience Measurement*. August 1990
- Syfret, Toby. *Television Peoplemeters in Europe*. WARC, London, 2001