SUMMARY

There are currently 21 peoplemeter panels operating in 18 countries in Europe. At a European industry level, a great deal of work has been done on the subject of harmonization of television research. The new comprehensive survey from EAAA was conducted over a period of 18 months by Dr. Toby Syfret, media research advisor to the EAAA. The report is invaluable in identifying opportunities to reduce variability and increase the unity between systems.

From an advertiser's point of view, some of the variations that were shown in the report appear significant. At a very basic level, a multi-national advertiser needs to know if a GRP in one country is the same as a GRP in another.

Concentrating on the methodological variations in the calculation and reporting of GRPs, Carat identified three areas for investigation: operational definitions, the components of a commercial rating (eg. guest viewing) and the definitions of reporting categories. In order to evaluate their effects, Carat commissioned RSMB to re-calculate the ratings for a selection of UK schedules using the respondent level data from the BARB panel.

This simulation demonstrates that the conventions adopted in some countries can lead to large variations in reported commercial impacts.

1, INTRODUCTION

Since last November, when the EC council of ministers reconstitued itself as the European Union, we have become citizens of a Europe united for the first time politically as well as commercially. The European Union was the culminating step in a process which has already harmonised trade, finance and some laws, and has virtually removed all internal quotas and tariffs.

In fact, a powerful impulse towards **integration** is at the root of nearly everything that Europeans have done for a generation. Integration has contributed greatly to European nations' wealth. In 35 years of ever-increasing unity, we have created a cohesive market of 350 million people, whose trade with each other increases by 9% each year.

We hear a lot now about the internationalisation of Europe's media business. All our countries' media market trends are converging. Trends like:

- **Fragmentation** of media audiences
- **Proliferation** of media vehicles
- **Concentration** of power in the hands of fewer, bigger media enterprises
- **Regulation** of the media

If this is true, it would be reasonable to assume that units of media measurement were also converging. Surely Europe should by now be getting close to a Single European Currency in media audience data, and a single standard for its application.

In an age of multi-national marketing, when products, packaging, manufacturing and distribution are being standardised across Europe, fundamental television differences between countries pose an increasing problem for the advertiser. We know that each market in Europe has different media structures, media usage and methods of reporting. What is not so obvious is how much of the observed differences in media usage relate to the media structure and national culture, and how much of the differences reported are simply a product of divergent research methodologies. Put very simply, is a Gross Rating Point in one country the same as a Gross Rating Point in another?

Carat has done a great deal of work already in this complex and challenging area. Right from the start, we have had to ask ourselves a fundamental question. Is comparability of television data really worth the trouble? Is TV data harmonisation only a preoccupation for a small group of giant multi-national advertisers.

The answer is: Yes, comparability is important, due to two key factors:

1. Fewer and fewer advertisers can remain concerned only with their home market, especially when they increasingly compete with big multi-nationals, and are subject to television campaigns spilling in from other territories.

2. Even more important is the enormous increase in television audience data being collected and processed across our continent. Those in charge of planning and executing development programmes, Europe-wide or country-by-country, will need centrally accessible data sources, giving comparable data on:

- Market potential
- Consumer attitudes and behaviour
- Competitive position
- And of course mediapenetration and performance

Business is no longer global *or* local. It is both at the same time.

Harmonisation of TV audience research has been a major industry preoccupation for over five years. In the past, different countries have measured TV audiences with different tools. Some countries were already using people-meters ten years ago, with the advantages of increased accuracy and faster data processing to determine detailed viewing patterns. Other countries continued with self-completion viewer diaries. Others again chose day-after or simultaneous telephone interview techniques. The results produced by these different methods were incompatible with each other.

More recently, a degree of harmony has emerged, at least on technical systems of data collection. People-meter technology has spread throughout Europe, and become an industry standard. There are 21 people-meter systems operating in 18 European countries. Added together, the panels of people-meter homes from which Europe's TV viewing is measured comprise just over 25,000 households, with 70,000 people reporting their viewing through handsets.

There has been an increasing technical standardisation of the **electronic equipment** with which viewing data was being gathered and compiled. Among the superficial similarities, however, Carat has seen significant differences from country to country in the **characteristics** of the data provided.

Industry sources, and in particular the European Association of Advertising Agencies, have already done a good job of tracking and comparing the basic specifications of each country's people-meter system. But this left many important questions unanswered: questions about the effects which different measurement methods and conventions will have on a client's advertising achievement figures.

As perhaps the leading user of people-meter data in Europe, Carat set up a working party to investigate, which drew on the experience of our European network of companies. Our aim was two-fold. First to isolate the differences in methodology and quantify their effects. Second, to determine which of these differences, if any, actually matter in real terms.

Studying each country's approach, the working group quickly identfied significant variations in the methodology by which commercial viewing is reported, after the minute by minute rating has been calculated.

The differences are found at the fundamental level of:

1. Operational Definitions: How a commercial TV rating is defined
2. Components of Ratings: What is actually being measured
3. Reporting Categories: How the audience to be reported on is classified

1. The first problem is: Determining which clock minute a commercial should be assigned to, in order to determine the size and type of audience it has achieved. (Commercials do not usually start conveniently at the exact beginning of a clock minute). How do we determine which clock minute each commercial belongs to, and therefore the audience it has achieved? There are seven possible answers, and different countries in Europe operate according to their own different criteria. Does it matter? We have examined specific commercial breaks and in just one example, the difference between the lowest and the highest estimates of a particular commercial's rating can mean a variation between a low of 7.0 and a high of 11.1. a difference of around 60%.

2. More than half of all European homes now have one or more VCR's. In some countries the figure is 60 or 70%. The decision to include or exclude "time-shift viewing" in the official ratings will make a difference to the end-result, and in particular for certain programme types such as movies.

Next, guest viewing. How do you treat visitors to a home, who sit down and watch TV? They are not part of the family whose TV habits are being measured by the people-meter. Yet they are part of the audience in the home being measured. Whether you ignore them or include them affects the ratings recorded. Most people-meters are equipped to register the viewing of up to eight additional viewers, beyond the basic household, although some can only deal with one or two. However, not all panels will include guest viewing in the reported commercial GRP.

Finally there is viewing during holiday time. When members of a people-meter panel go away on holiday, do you keep them in the panel or not? This is not an academic question. If they stay in the panel, they are assumed to be available to view; but of course they don't, because they are away on a beach somewhere. Ratings accordingly go down. If you temporarily throw them out of the panel for the duration of their holiday, the ratings will go up again. This possibly allows for any viewing they may be doing at their holiday destination. But does this mean ratings are now being inflated? People really are absent, after all. Different countries see the problem in different ways.

3. The third and last type of variant our working party needed to consider was reporting categories. Across Europe, the minimum age of a people meter panelist can range from 3 to 6 years old, thus affecting the definition of an Individual GRP, and there is no standard age definition of an adult, and there are five different definitions for a housewife/housekeeper. Is this just academic? It could make a difference to Adult or Housewife TV ratings if the **basic definitions** of adults and housewives are fundamentally different.

In total there are over **50 different possible variants** which have to be taken into account, any one of which can be an obstacle to harmonising the audience figures for TV schedules across Europe. Taken together, they mean that we are faced with a measurement task in which our only available measure is "elastic".

However, there was another important question we had to ask. Does all this really matter? What difference does it make to the results of an advertising campaign? To answer that question, the Carat Working Party approached RSMB to examine a series of **real TV advertising campaigns,** already transmitted and evaluated in their home country, and then to re-calculate the ratings for each individual spot in each schedule, according to the methods and classifications in use in each European country.

2. THE EXPERIMENT

The UK BARB audience measurement panel is actually operated as 17 separate panels, each representing a different ITV area. For this experiment we chose the Midlands ITV area panel - this has about 1400 individuals and is the largest. Carat selected nine schedules for evaluation. These had a variety of target audiences, a mixture of spot lengths and different distributions by day-part. Each was a mixture of ITV and channel 4 spots.

The objective of the first part of the experiment was to evaluate the **marginal** effects of the different operational definitions, the various components of the ratings and the different reporting category definitions.

Before moving onto the results, it is important to consider the BARB definition of viewing in a minute which we accepted as the base for this experiment:

1. In the UK, TV viewing is defined by a person's presence in the room where the TV set is on.

2. TV set and person viewing records are edited separately. In each we start with the actual record and move through two editing steps to the minute viewing records.

3. In the first step, anything which lasts for less than fifteen seconds is ignored by the meter. This is the persistence level used in the UK.

4. In the second step, the status that persists for the majority of a particular minute is assigned to the whole minute. For example, if the TV set was tuned to ITV for forty seconds, then the minute viewing record is ITV for the whole minute.

So the database for this analysis was a minute by minute viewing record for every panel member. This was used in conjunction with the commercials log which has the start time and duration of every commercial timed to the nearest second.

In France and Portugal, viewing is reported to the nearest second rather than to the nearest minute. We have estimated the second by second ratings by assuming that the BARB database measures the rating at the centre of each minute and then drawing a straight line between adjacent points. We recognise that this cannot take account of all the features of second by second data but at least there is a measurement of the gradient.

3. RESULTS

The results are covered in three sections, starting with the various **operational definitions.** The data shown in table 3.1 are for all individuals aged four plus and schedule GRP's based upon live viewing including guests.

For each of the nine schedules, we have calculated the total GRP's using each of the nine operational definitions which exist. These have then been compared with the UK definition which uses the rating of the minute in which the spot started. This definition is also used in Spain, Sweden, Denmark, Finland and Norway.

The first column in table 3.1 shows the average percentage difference for the nine schedules. The second column shows the range in the percentage differences across the nine schedules.

Table 3.1

EFFECT ON SCHEDULE GRPs: OPERATIONAL DEFINITIONS

Live + Guests		
Individuals Age 4+	***% Difference***	
	Average	**Range**
Spot Start Minute Rating	**0**	**0,0**
Spot Majority Minute Rating	-1	-3,0
Weighted Average of Spot Minutes	-1	-2,0
Spot Average Second rating	0	-2,0
Break Average - Minutes	0	-2,+l
Break Average - Seconds	-1	-3,0
5 Minute Average rating	0	-2,+2
Break Reach	+ 12	+6,+16
15 Minute Reach	+50	+24,+61

The Netherlands report the GRP of the minute in which the majority of the spot falls. Italy and AGB Portugal report an average of the GRPs of all the minutes into which the spot falls, weighted according to the proportion of the spot in each minute. Ecotel Portugal report the GRP for the actual duration of the commercial in seconds. These definitions might all be considered to be more representative of the actual commercial ratings and in each case there is a small reduction in the schedule GRPs.

Austria, Belgium and Germany report the average GRP of all the minutes in the commercial break. Ireland report the average GRP of the five minutes from the beginning of the break. These definitions do not produce GRPs which are systematically different to the UK definition, although this does vary according to the positioning of each schedule's spots within the breaks.

France reports the average GRP of all the **seconds** in the commercial break, giving slightly lower GRPs.

The greatest variation was in the two Swiss definitions - the "break reach" and the "15 minute reach" planning definitions. This is obvious but it is useful to quantify the differences and to note how much they vary across the nine schedules. However, we must sound a note of caution here because these results are based upon the UK broadcasting environment and audience behaviour. It may be that the Swiss audience is more consistent meaning that spot ratings are closer to break reach than in the UK.

The evaluation of **components of ratings** was based upon schedule GRP's for individuals age 4+ and the UK's "spot start minute" definition. The results are shown in Table 3.2.

Table 3.2

EFFECT ON SCHEDULE GRPs: COMPONENTS OF RATINGS

Spot Start Minute Rating		
Individuals Age 4+	***% Difference***	
	Average	**Range**
Live Viewing Only	**0**	**0,0**
Including Guests	+6	+2,+9
Including Timeshift	4-2	0,4-3
Excluding Holiday Homes	+3	4-3,4-3

Guest viewing was the most important component, generally adding between six and seven percent. The exceptions were the schedule at 9% - this had more late night spots - and the schedule at 2% - this was confined to breakfast time.

On average, the inclusion of viewing to timeshift commercials added two percent to the GRP's. But there was no increase for the Konica breakfast time only schedule, but then we would not expect time-shift viewing of UK breakfast time TV.

The holiday effect was a constant 3% for all schedules, which would be expected as we simply removed all holiday takers from the base, thereby eliminating a number of nil viewers. However, this analysis is weak because we don't know which BARB panel members were actually on holiday. We simply checked for homes with 7 day's nil viewing in each of the four weeks we analysed. This would miss people whose holidays start and end in mid-week or cases where some family members stay at home.

However, we know that the holiday effect will relate exactly to the percentage of the population on holiday and will therefore be highest in certain months and affect different countries in different ways. France, Portugal and Turkey exclude holiday homes.

The evaluation of **reporting categories** was for live viewing including guests and based upon the UK's "spot start minute" definition. The analysis has so far been confined to differences in the definitions of All Individuals, Adults and Housewives. The results are shown in Table 3.3.

Table 33

EFFECT ON SCHEDULE GRPs: REPORTING CATEGORIES

Spot Start Minute Rating		
Live + Guests	*% Difference*	
	Average	**Range**
Individuals 4+	**0**	**0,0**
5+	0	-3,+l
6+	+ 1	-6,+3
6-74	0	-5,+2
Adults 16+	**0**	**0,0**
12+	-2	-3,0
14+	-1	-2,-1
15+	0	0,0
18+	+ 1	0,+l
15-74	-2	-4,+l
Housewives MF16+	**0**	**0,0**
MFI 8+	0	0,0
F16+	+3	+2,+4
F16+ not working	+25	+ l,+32
All F15+	+2	-5,+7

Generally individuals aged 5+ and 6+ will watch slightly more than individuals aged 4+ - the sole exception here is again the breakfast time only schedule.

As the age break for defining adults gets older, so the GRP's get higher. Compared with age 16+, the variation was from -2% at age 12+ to +1% at age 18+. Eliminating older viewers - as the Swiss system does - results in small decreases in the ratings. These will be more or less extreme depending upon the day-parts used. For example, for one schedule the GRP's dropped by 4% reflecting it's day-time bias.

In the UK we have one housewife per household who must be age 16+ but can be male or female. Restricting housewives to females increased the GRP's by about 3%. The Greek definition, which specifies that a housewife must be female and have no other occupation outside of the household, increases the GRP's by an average of 25%; the exception is the breakfast time only schedule.

In Norway and Sweden, the concept of a housewife is not recognised, therefore by default we have considered all adult women. The effect varies according to the schedule structure, showing a 5% decrease for the breakfast time schedule and a highest 7% increase for the schedule with a night time bias.

When considering reporting categories, it is also important to note the variations in the universes which combine with ratings to generate overall variations in the thousands viewing. Table 3.4 demonstrates that the effects on universes are much greater than the effects on GRPs.

Table 3.4

EFFECT ON UNIVERSE; REPORTING CATEGORIES	
	% Difference
Individuals 4+	**0**
5+	-2
6+	-4
6-74	-9
Adults 16+	**0**
12+	+7
14+	+3
15+	+ 1
18+	-3
15-74	-5
Housewives MFI6+	**0**
MFI 8+	0
F16+	-9
F16+ not working	-54
All F15+	+18

4. COMMERCIAL IMPACTS

This leads us on to the consideration of commercial impacts and the second part of the experiment. Each country has a different specification in terms of the operational definitions, components of ratings and reporting categories. Each different specification was applied to the UK BARB database in the calculation of total impacts for each schedule.

Based upon an average of the nine schedules and compared with the UK currency, estimated impacts varied considerably between countries:

Individuals:	from -8% to +28%
Adults:	from -7% to +31%
Housewives:	from -46% to +23%

This demonstrates how misleading simple cost per thousand comparisons can be.

5. REACH AND FREQUENCY

The other key analysis used in the evaluation of schedule performance is reach and frequency. It is not obvious to us how some of the operational definitions fit in with reach and frequency analysis, so we have confined our investigation to the UK definition of viewing to a commercial spot.

Many countries base reach and frequency analysis on the continuously reporting panel for the duration of the schedule. This generates a complete record of the spots viewed for each panel member.

Use of a non-continuous panel is justified when panel continuity is high and daily non-reporting is low. Reach is not affected much because the chances are small that the only day that a person saw the campaign is also a non-reporting day.

In the UK, we base reach and frequency analysis on the sample which reports on the middle day of the schedule. Non-reporting on other days is assumed to be nil viewing. The added advantage of this approach is that we can use the processing weights already calculated for that day: The effective sample is representative. We can even compensate for the small effect of non-reporting by using a probability model to gross-up to the published GRP's. This generates an appealing level of consistency and also allows us to incorporate guest viewing into our estimates of schedule reach. In this context, we are implicitly assuming that guest viewing is a surrogate measurement of a panel member's out of home viewing.

Table 5.1 shows the effects on schedule reach of including guests and using non-continuous panels. The continuous panel estimate of schedule reach has been taken as the base for comparison at different levels of schedule GRP's.

Table 5.1

REACH AND FREQUENCY

Schedule	*% Change in Schedule Reach From:*	
GRPs	**Guests at 5%**	**Daily Non-Response at 2%**
100	+2.5%	-1.0%
200	+ 1.6%	-0.7%
300	+ 1.2%	-0.5%
400	+0.9%	-0.4%
500	+0.7%	-0.3%
600	+0.6%	-0.3%

The data for guests show the percentage increment in schedule reach resulting from a 5% increase in GRP's from adding guests. So at 100 GRP's, guests would increase the schedule reach by *2\%:* In the UK this would be an increase in reach from about 50% to 51%. As the GRP's increase, the percentage increase in reach diminishes because we move towards a saturation level.

The data for the non-continuous panel show the percentage loss in schedule reach resulting from the use of the non-continuous panel with a daily non-response rate of 2%. At 100 GRP's the loss is 1% but this diminishes to \% at 600 TVR's.

6. CONCLUSIONS

Although the differences between peoplemeter systems appear to be significant, not all of them are. The most important operational variation is the use of break reach to estimate spot ratings.

Guest viewing, timeshift and holiday homes all have an obvious effect. The definitions of individuals, adults and housewives have a greater effect on impacts than GRPs.

The results show that many of the effects are dependent upon the structure of the schedules.

It is clear that the conventions adopted in some countries can lead to large variations in impacts and it is vital to have an understanding of these differences.

REFERENCES

Television Peoplemeters in Europe, EAAA, European Association of Advertising Agencies, June 1993.