ESOMAR / ARF Worldwide
Electronic and Broadcast Audience Research Symposium

Electronic Media and Measurement Trends – on a Collision Course?

Vienna, Austria
26th – 28th April 1998

Measuring Small Audiences:
The Challenge to Audience Measurement Systems

by: Tony Twyman and Steve Wilcox

The paper identifies the different ways in which media developments lead to ever smaller audiences being examined within peoplemeter systems.
This consequent unreliability of much peoplemeter data is assessed and strategies proposed for dealing with this. A greater use of aggregation in the use of audiences and trading practices is proposed. Alternative techniques are also considered.
A key principle is that the measurement of a number of small audiences in aggregate can be as reliable as the measurement of a large audience. Not only does research have to change but also the use of research.

1. Introduction

All developments in electronic and broadcast media leads towards more stations, more choice, more targeting and the fragmentation of audiences. The creation of yet more small audience channels does not, however, eliminate the appeal of mass audience channels. These remain as mass audience advertising media, with a continuing demand for the spot by spot assessment of campaigns which has been a key feature of peoplemeter measurement systems.

The newer, smaller, media are likely to sell their advertising in a radically different way from the mass media, with packages of spots, even packages of stations, replacing the old single spot unit of advertising measurement. They will, however, often be selling to advertisers using the mass media and wanting comparability between the meanings of assessments for their different campaigns.

The challenge to the current style of national measurements systems is how to accommodate the ever widening range of audience sizes which it is expected to measure. This is a feature of a recent paper presented by Read & Johnson (1997) in which they discuss the development of the next UK audience measurement specification. The core of the problem is that the smaller the audience the larger the **relative** size of sampling error. This implies potential increases in sample sizes of a scale which exceeds the likely expansion of advertising revenue and research funding available. Paralleling the diverging advertising demands on research systems, the broadcasting programme makers also, will have different requirements, according to the varying nature of their programming.

It is important to recognise, however, that there is not just one ‘small audience problem’ but a number with different potential research solutions and even in some cases no conventional solution.

2. The Measurement of Sampling Error

Key to any discussion of the measurement of small audiences is a realistic appreciation of the extent of sampling errors involved. We will be referring to and summarising extracts from work in the UK and elsewhere.

First, however, we must make clear what we mean by sampling error.

In the purest sense the term sampling error is used to mean the deviation of behaviour of a randomly selected sub-sample with no response bias, from the behaviour of the total population. Practical situations are far from this. Samples are not purely random and there is response bias but we still need to understand the variability of the data.

In using the term ‘sampling error’ here, we want to express the degree of variability which any measurement or comparison between measurements is subject to when there is no real change in the behaviour which it is intended to measure.

With panels there are a number of factors which contribute to the amount of statistical variability:

1) When panels are initially recruited, the sample will be biased through:
   a) differential non-response biases inherent in the system.
   b) chance features of the particular sample selected which remain as a sample bias, changing over time as panel membership changes.
2) Comparisons between measurements at different times involves many of the same individuals and according to the degree of correlation between their behaviour at those times, there is a reduction in sampling error. This correlation diminishes over time however as people get older, change their social life, their work life and their interests.
3) The need to balance known demographic imbalances involves weighting which can significantly decrease effective sample size and increase sampling error.

3. Types of Small Audience Situations

In this paper we seek to identify the range of small audience situations, the likely data requirements and how they might be researched.

We are not starting with a 'clean slate' here. In most countries with broadcasting systems sufficiently developed to generate small station problems there will already be sophisticated peoplemeter systems.

These have been designed initially to measure mass audiences but have been progressively expanded and adapted to report on smaller audiences. Research solutions have to be considered:

- within existing systems
- by expanding and adapting existing systems
- by creating entirely new research sources

3.1 Viewing By Smaller Sub-Groups To Larger Stations

This is essentially a problem of sample size.

This is a situation which regularly occurs within existing systems and even for mass audience channels.

There appears to be a law whereby, whatever the sample size, the number of sub-groups reported expands to include many for which the sample size is inadequate.

Within the current BARB system the sample sizes and sampling errors shown in table 1 for the largest regional panel illustrate the point.

Insert Table 1 here

These sampling errors assume the panel to be perfectly balanced. In reality the sampling errors are up to 20% larger for the actual panel which is weighted to correct for demographic profile imbalances. So these sampling errors represent the best that could be achieved given the total number of homes available. It is salutary to note that these sampling errors are for a peak time rating on the largest commercial TV station in the UK.

Any attempt at optimising the choice of individual spots on the smaller sub-groups is clearly a waste of time.

One common response to statistics such as these is that practices should change and that trading should be based on larger more reliable sub-groups and/or that optimisations and appraisal should be in terms of schedules rather than individual spots. The statistical reliability of this approach is discussed in section 4.

Another approach under test in the UK for regional sub-groups is that of modelling or factoring from the network panel.

The principle is that a regional panel measures the main audience categories directly. Sub-group audiences are then factored by applying the relationship between the sub-group and the main category found at that time on the larger network panel, to the directly measured regional main category audience. This factor is derived after weighting the network data to match the demographic profile of the region. A summary of this is provided in section 6.

This approach can be used to reduce sub-group variability equivalent to increasing effective sample size by between 50% and 100%. This is an increase beyond the levels of affordability, but even that is not enough where the market is trying to trade on sub-groups with samples of 50.

This approach we believe could be used in the UK and help to make sub-group data more reliable. It is, however, viable only in conditions like in the UK where there is a broad framework of consistency in programming across the regions within the network and few marked deviations of programming style at the regional level. We have not found evidence that regional variations in programming do affect the validity of factoring.

Factoring however applies most readily to a regional panel structure for mass audiences. It is no solution for small area channels like cable or niche channels.

Overall our solution to small sub-group audiences on larger panels is to suggest that the sampling errors should be examined (as above), samples increased to what is affordable and to accept that trading on spots for those sub-groups which cannot be measured reliably will be unproductive.

Steps which can help are trading and appraisal in terms of schedules of spots and in appropriate cases, factoring.

3.2 Viewing By Large Sub-Groups To Smaller Stations Or Large Stations At Off-Peak Times

Put more simply this is the problem of small ratings on large panels.

As competition increases, audiences fragment and there are always:-

- stations which always have low ratings.
- times when even large stations have low ratings.

This situation occurs increasingly within panels designed to measure mass audiences.

Where stations are restricted by access such as for satellite or cable, panels representative of those sub-sections of the universe can be recruited.

In the UK, homes with satellite or cable are broken out of the main panel and weighted as a **Network Satellite Panel.** This provides around 1200 households and 3600 individuals, without any additional boosting.

Such is the fragmentation within these homes however, that many stations record permanently low weekly audiences. These data are robust within satellite homes for the terrestrial channels, for Total Sky and other aggregations of channels. Some channels however regularly record an average of 1 or 2 minutes of viewing per head each week.

In a sense these figures are reliable in that they always show very low audiences week after week.

Where the problems arise, for all channels, is when individuals spots or programmes are considered. For many of the larger satellite stations, even within the satellite universe, many ratings at individual times are 1% or less, often 0.1% or less.

The sampling errors on these are enormous. For example, consider the largest satellite channel in the UK. Amongst all housewives this channel took a 4.5% share of all viewing in satellite receiving homes in a recent week (w/e 25th January 1998). (Note that the next largest satellite channel took only a 2.6% share.) In this particular week, two thirds of this channel's programmes had housewife ratings of 1% or less and one third had housewife ratings of 0.1% or less. The 95% confidence intervals on housewife ratings of 1% and 0.1% in satellite homes are ±60% and ±180% respectively. (Again these sampling errors assume the panel to be perfectly balanced; in reality they are larger.)

Two thirds of the satellite channels reported by BARB never achieved any rating as high as 1% in this particular week.

If audiences are expressed in terms of numbers of viewers, however, they take on a reality which belies their statistical bias. For example, a programme with a rating of 0.5% could easily lose all its audience from one week to the next purely as a result of sampling error. Amongst housewives in satellite homes (there are 6.6 million of them in the population) this represents a drop in the audience from over 30,000 to nothing at all. How can you lose 30,000 viewers from one week to the next?

Because they represent small ratings with big sampling errors the variability looks implausible.

What is the solution? This depends on the purpose for which audience research data are needed.

Programming: It is certainly possible to see which are the most successful programmes even from highly unstable small ratings data for small stations. Judgement is considerably improved by managing several weeks data together.

The precision of assessment for programme audiences is less than for large rating channels but the need for more subtle distinctions may be less. The differences between programmes may even become more deviant because they may be less affected by competition.

Buying And Selling Advertising: Here any attempt to work on an individual spot time may be a waste of time. Improved assessments may be made by:

- using data averaged over time.
- assessing whole schedules either within a channel but more realistically across a number of channels.

For advertisers on niche channels representing a special market eg. a computer channel, advertisers may well wish to get an idea of when the best times to advertise are. They are however more likely to buy a schedule and compare the direct response with campaigns in other media. It is possible that the more specialised the market the less precise audience estimates are required.

3.3 Viewing To Stations With Restricted Universes

With panels covering the whole television universe, some restricted universes may be represented with only small sample sizes and there may be difficulties in representing their characteristics.

Restricted universes in this sense occur in a number of ways.

3.3.1 Limited Regional Coverage

In the UK some cable franchises have small catchment areas. Within BARB, cable as a whole is represented by a panel which is a specially weighted sub-set of the main panels. This reports separately on cable stations which have a wide geographical coverage. Small regional cable franchises may, however, wish to know the patterns of viewing to the station mix which they offer and their own local cable services. Their coverage by a network peoplemeter sample is negligible. It would be possible but not economically viable to recruit a special peoplemeter panel for the area. Instead in the UK the Cable Research Group have commissioned, outside BARB, periodic two-week paper diary studies using diary formats not unlike those used for much radio research. Some of this work is described in section 5.2.

This kind of situation is likely to increase for the future.

Most regional television structures end up with regions that vary in size. This often means that the smaller stations would not have an adequate sample based upon proportionate regional sampling. The solution is usually disproportionate geographical sampling or a federation of regional panels.

Whilst strict statistical logic would demand equal sized panels everywhere the money at risk argument often leads to compromise whereby larger areas may be capped off at a certain limit and smaller areas boosted up. The UK is an example of this illustrated by 3 of the 13 regions areas:

	% Of National Population	% Of Meters	Sample Size Households
London	20.2%	11.7%	525
North East	5.3%	6.1%	275
Border	1.2%	2.2%	100

One of the problems with this is that there is sometimes a tendency to treat all the areas as having the same currency available. Thus the sample size for Border hardly warrants pursuit of spot by spot buying, certainly not for sub-groups, but it sometimes happens.

Possible remedies include selling by schedules or aggregated ratings and factoring discussed elsewhere.

An example of equal sized regional panels is Belgium with two equal panels of 750 Households for each of the Flemish and French speaking parts. Paradoxically although the national sample is about a third of the UK's, the actual panels used for trading are bigger. A regional programming and trading structure is one situation where, even with mass audience channels, the use of peoplemeter panels leads to statistical strain. It seems likely, however, that there is a general trend towards trading television advertising in larger units which may ease this.

3.3.2 Services Based Upon New Technology

The advent of satellite transmissions was a past example of this. There it was possible to recruit a special sample of satellite receivers who had a vastly increased range of programme choice compared with terrestrial reception. In the UK there was initially a separate panel but ultimately a specially weighted sub-sample of the main peoplemeter panels was used. This currently provides a sample of around 1200 households.

We expect that digital television will be measured in the same way in the UK. This could initially involve special extra panels for satellite digital, terrestrial digital and even digital cable services since there is very likely to be much initial overlap between this mode of reception. Such panels could be merged with analogous panels when the universe is large enough.

These new developments present special small audience problems:

a) **Universes**

It is easy to define access to equipment but harder to measure it when it starts from zero and may rise rapidly and erratically. Within the broadest definition of having the reception equipment however, there is the added complication of subscription packages incorporating different channels. These are subject to an additional variability from take-up and churn within the variability of the equipment universe.

Universes have been generally obtained from some independent survey source. An Establishment Survey for example, for a slowly changing terrestrial source can provide:-

- a reliable estimate of universe size.
- a profile of demographic and often characteristics of the universe.
- a source of households for panel recruitment.

With new services such as digital transmissions the new problems are that the universes are:-

- initially very small.
- dispersed through the population.
- changing very rapidly; for individual stations up and down.
- highly complex in terms of combinations of channels received.

These characteristics mean that no representative sample is likely to be large enough, affordable on a continuous basis nor even able to be processed quickly enough.

This means that some alternative approach is necessary. In practice broadcasters will have exact databases showing who is paying for and receiving what on a near-daily basis. It would be logical to use this. The objection is sometimes raised that broadcasters might inflate the figures and/or be able to detect the identity of the panel home. It will be necessary to counter this by some form of independent auditing and access to the database. It will also be necessary to create new legal safeguards and protection against interference with panel homes. The use of a broadcaster's database does not uniquely create this problem, it is there from the moment that the broadcaster gets into a direct one-to-one on-going relationship with the households in the audience. Ultimately intelligent digital decoders will be able to record station viewing data in great detail on large samples. This will need some individual viewing data from smaller samples modelled onto it. This would solve both the universe and the small audiences problem for marketing strategies based on type of household rather than type of individual.

b) **Audience Fragmentation**

New technologies bringing more choice and greater fragmentation. Digital television is likely to extend the range of channels from the 30+ of satellite into the hundreds. Different channels will show the same films at different times to provide a near video-on-demand service. Necessarily, most audiences will be very small, a further extension of the issues discussed in section 3.2. This means that for all but a few channels the assessment of individual spot ratings will be pointless. We would expect to see television planning and assessment based upon aggregated data probably involving selling of schedules comprising many small ratings spread across a range of channels.

Once again the receiving of schedule audience measurement becomes of crucial importance.

c) **Panel Structure**

As services develop those which can afford peoplemeter panels will probably do so but initially, with relatively small sample sizes. It will also be necessary to control panel membership in terms of:-

- combinations of channels received.
- novelty effects ie. length of ownership.

This will require complex weighting reducing effective sample sizes even more and exacerbating the problems of fragmentation discussed above.

Only data aggregated across channels and/or times will be robust.

When choice gets this complex and with the development of electronic programme guides where programmes can be chosen without channel awareness, the option of using alternative techniques such as paper diaries and recall will no longer exist.

The industry will therefore have to get used to using audience measurement data for small audiences from small peoplemeter panels in a responsible way using aggregated data. Until that is the intelligent decoder is able to give precise set range data on large samples.

Programming needs will vary according to the nature of the channel. Even with very small share channels it is possible to see which are the most popular programmes particularly if schedules are consistent and weeks averaged together. BARB currently publishes Programme Top Tens for many small share channels, which are robust, in the sense of similar programmes appearing week after week. Any subtlety in terms of small differences between audiences would however, be impossible. Programmers wishing to fine tune programmes or schedules would probably gain more from qualitative research among viewers to their programmes.

3.3.3 Ethnic or Language Minorities

Ethnic or minority language groups are likely, by definition, to have low representation on general representative samples. There are likely to be special sampling problems in that such groups are both clustered but not exclusively confined within any geographical boundary. Universe measurement and sample selection probably requires large scale surveys and some allowance made for differential non response. Even then the sample may not be adequate to provide reliable data for channels servicing these groups.

One solution is a separate peoplemeter panel. This occurs in the UK for Welsh Speakers, measuring audiences to S4C. There is no separate panel for Gaelic speakers in Scotland. Response to programmes in Gaelic is studied through qualitative audience appreciation studies. The Gaelic channel in the Republic of Ireland has Gaelic speakers as a possible audience sub-group on the main panel. In Germany foreigners have been excluded from the main television panel universes but may now be represented by a separate panel.

Whether an ethnic minority or language channel has a separate panel is largely a matter of economics.

Cable and the development of digital services will make niche channels possible for smaller ethnic groups. The limited data available from mainstream panels may mean that alternative techniques have to be used.

3.3.4 Viewing to Minority Interest Stations, Intermittent Interest Channels

The multiplication of choice will give rise to channels which have a very restricted 'niche' appeal but one which is not identifiable by region, language or ownership of equipment. The channel would be based upon interest in a topic such as natural history, or history. An intermittent interest channel would be a weather or traffic channel. These stations are essentially general in potential appeal but likely to achieve a low reach and share. They suffer from the general small channel problems and the solutions lie in aggregation as already discussed.

For minority stations where there is a marked minority appeal, there may be problems not only of sample size but also of panel bias. The chance over or under representation of a minority interest group could stay with the panel for some time. Here alternative techniques with larger independent samples may help (see section 5).

Section 4 - Sampling Errors for Small Audience Measurements

4.1 Small Audience Measurements

In the UK the BARB TV peoplemeter system currently reports audiences to five national terrestrial channels, one local terrestrial channel (S4C in Wales), 38 channels delivered by satellite and cable (this number is constantly changing) and five cable exclusive channels. Typically, the national terrestrial channels account for the following audience shares:

BBC1	30%
BBC2	11%
ITV	33%
Channel 4	10%
Channel 5	3%

Of the 38 satellite/cable channels, only two account for more than 1% of all viewing. The full distribution is as follows:

Share	**Number of Channels**
1-1½%	2

½-1%	6
0-½%	30

In total, the five cable exclusive channels account for less than ½% of all viewing, as does the local terrestrial channel S4C.

A large number of cable exclusive channels are not reported by the BARB system because the data is not considered to be sufficiently robust. (These are catered for outside the BARB system see section 5.) The arrival of digital TV later this year will generate yet another small audience measurement requirement.

The small channel shares are partly due to the large numbers of channels available and partly because only 34% of the population have access to cable or satellite and only 13% are cable exclusive.

In the UK, the national terrestrial channels are also commonly reported on a regional basis, in terms of either the 12 BBC editorial regions or 16 ITV areas. This is another key dimension resulting in small audiences. For example, if 10% of the population live in a particular ITV area, then the share of all TV viewing by the whole national population which is accounted for by viewing to the ITV station broadcasting in that area is only 3.3% (i.e. 10% of the national ITV share of 33%). Effectively this is another example of a restricted availability channel because only 10% of the population has access to that particular regional ITV station.

The last dimension which results in audience fragmentation is the need to report on demographic categories, ranging from simple Male/Female splits to very tightly defined age groups. For example, a 10% penetration sub-group's viewing to an ITV station in an area containing 10% of the population would only account for 0.33% of the total national populations' viewing.

Of course it is not normal practice to report such fragmented audiences as percentages of the total national population base. Therefore the percentages are not normally seen as such small numbers. However, this way of presenting the audiences is a useful lead in to the consideration of sampling errors and the relative reliability of the various audience measurements.

4.2 Sampling Error Study

BARB and RSMB have recently completed the first phase of a major study of the sampling errors associated with the various audience measurements produced by the TV peoplemeter panel in the UK. This is considered to be an essential contribution to the sample design component of the future audience measurement specification. The theory has been developed to allow the calculation of sampling errors for many different audience measurements and to compare the performance of perfectly balanced proportionate and disproportionate designs and to assess the effect of weighting used to correct for the usual panel imbalances that exist within an operational system.

4.2.1 The Calculation of Sampling Error

Several papers have been written concerning the components of sampling error and the methodology for their calculation (eg. Schillmoeller, 1992, Boon, 1994 and Twyman & Wilcox, 1996).

The calculation of sampling error takes account of the variability in the audience measurement between individuals, the sample size, clustering within households and weighting. These factors and their effects can be different for each measurement, for each channel, for each demographic category and each area base.

It was necessary to consider the whole range of different audience measurements because some will have smaller sampling errors than others and therefore may be more useful in the small audience situation:

- Average ratings and channel share for all-time, time segments (day-parts), quarter hours and individual minutes.

- Channel reach.
- Programme, commercial break and individual commercial spot ratings.
- Reach and Frequency analysis.
- Daily, weekly and four week averages.
- Change over time, from month to month and from year to year.

The actual analyses were based on a limited number of channels using two ITV area panels. For the purposes of this paper, the results have been interpreted to provide approximate sampling errors for a number of hypothetical situations.

All sampling errors have been converted to 95% confidence intervals. This means there is a 5% chance that the audience measurement estimate is more than one confidence interval from the "true" value of the audience measurement.

4.2.2 Sampling Errors for Proportionate Panel Designs

Because network based channel share encapsulates the extent of each small audience situation, a useful start point is to consider the sampling errors on channel shares for the types of viewing situations described in section 4.1.

The UK BARB panel is currently nearly 4500 homes with 8600 adults. If this were to be of a proportionate design and perfectly balanced (ie. no weighting were required) then the shares of viewing would have the following sampling errors. These are shown for a single minute, a single day and then for a four week average in table 2.

Insert Table 2 here

* Sampling errors have not yet been calculated for Channel 5, S4C nor any of the cable channels so interpolations have been made. It should also be noted that sampling errors have not been calculated for any satellite channels with less than 0.1% share. The single minute sampling error relates to peak-time.

At the next level of fragmentation (table 3), consider sampling errors for an average 10% penetration demographic sub-group or a 10% penetration geographical region. The network based channel shares are all divided by 10. The sampling errors above increase as the sample size decreases - ie. multiply by $\sqrt{10} = 3.2$.

Insert Table 3 here

For highly targeted channels this table must be interpreted carefully. This is because it would not be normal practice to analyse "average" demographic sub-groups. More often than not the key target sub-group would account for a large proportion of the channel's total audience. In this situation the percentage sampling error will not increase as much as the decrease in sample size suggests because the sub-group has higher viewing levels. In fact we can hypothesise that if we do have a situation where the whole of a channel's audience is attributed to one key demographic sub-group (eg. 16-34's and a "young" music channel) then the percentage sampling error for the sub-group is the same as for All Adults.

This can be demonstrated theoretically, using single minute ratings rather than channel share in order to keep things simple:

Network sample = 8600 adults

Single minute TVR = 1

$$\text{Sampling error} = \sqrt{\frac{pxq}{n}} = \sqrt{\frac{1x99}{8600}} = 0.107$$

95% confidence interval = ±0.214 = ±21%

Now suppose that the whole of this audience is contained within a 10% penetration sub-group. Then:

Sub-group sample = 860 adults

Single minute TVR = 10

Sampling error = $\sqrt{\frac{pxq}{n}} = \sqrt{\frac{10x90}{860}} = 1.02$

95% confidence interval = ±2.04 = ±20%

which is almost exactly the same and therefore almost totally independent of the sample size.

Empirical evidence for alternative audience measurements and with a more sophisticated sampling error calculation is not always so consistent. However the relationship seems to be proved in terms of orders of magnitude - certainly the sub-group would not have a sampling error 3.2 (= $\sqrt{10}$) times as large.

In order to complete the series of audience share sampling error tables, table 4 is for an "average" 10% penetration demographic sub-group within a 10% penetration region:

Insert Table 4 here

The original network based channel shares are all divided by 100 and the confidence intervals are now ten times as big as those shown in the national/all adults table.

4.2.3 Individual Spot Ratings vs. Schedule Averages

Having used channel share and "average" demographic sub-groups to demonstrate in principle how large sampling errors can be in small audience situations, it is important now to consider "real" demographic sub-groups and the key audience measurements used in the buying and selling of advertising. "Real" demographic sub-groups do not have average levels of variability nor average levels of clustering within households. The key audience measurements relate to individual commercial spots and whole advertising schedules.

First consider the sampling errors for the ratings to a selection of individual minutes broadcast on ITV and Channel 4, shown in table 5. The sample base is the London ITV area panel which comprised 530 homes, delivering 459 Men but only 62 Men 16-24. The analysis period is November 1996. This illustrates the small audience measurement situations arising from restricted areas, small demographic groups and times of low viewing.

Insert Table 5 Here

All the sampling errors are large, even for the peak-time ITV All Men rating. The Men 16-24 sampling errors are huge. The zero Channel 4 ratings actually emphasise the small sample problem - even a Men 16-24 rating of 5% (as achieved within All Men) would be the result of only 3 individual panel members viewing.

By averaging over time, even within a continuous panel, there will be significant reductions in sampling error. This fundamental theory originally expounded in a report prepared by Arbitron (1974) has been demonstrated in several published papers (eg. Wilcox & Reeve, 1992). For example, average ratings over four consecutive Mondays have the sampling errors shown in table 6.

Insert Table 6 here

Although there is some variability in the relationship with the single minute rating sampling errors - to be expected with real data and small samples - on average the percentage sampling errors are halved. We believe that many broadcasters are already using such averages for planning purposes.

This principle can be extended to whole schedules where in general we will find even greater reductions in sampling errors. Table 7 shows results for five schedules broadcast in November 1996, again based upon the London panel.

Insert Table 7 here

For All Men and for any schedule with a reasonable number of ratings, the sampling errors have reduced to a more manageable level. However, there does seem to be a plateau beyond which additional ratings will not result in further sampling error reductions. For All Men the minimum 95% confidence interval seems to be ±5% whilst for Men 16-24 it is about ±15%. This is approximately in line with their relative sample sizes although Men 16-24 are also more variable as a group.

However, the basic principle is clearly demonstrated: schedule total ratings have much smaller sampling errors than do individual commercial spot ratings.

4.2.4 Sampling Errors for Different Schedule Structures

A key question we have asked is how much the sampling error for schedule total ratings is dependent upon the composition of the schedule. ie. Is the schedule total ratings percentage sampling error high if the individual spots in the schedule have low ratings and therefore high percentage sampling errors? For example, is the sampling error for a schedule of 10 spots with an average rating of 20% the same as for a schedule of 20 spots with an average rating of 10%? The principle can be demonstrated with some simplistic theory.

For a single minute with a rating of p and given a sample size of n:

$$\text{Sampling error} = \sqrt{\frac{p \times (100 - p)}{n}}$$

For a schedule of s spots each with a rating of p (and assuming statistical independence between spots) then:

$$\text{Sampling error} = \sqrt{\frac{s \times p \times (100 - p)}{n}}$$

95% Confidence Interval = 2 x Sampling Error

Table 8 shows results for schedules of varying compositions but all delivering 200 ratings based upon the London sample of 459 Men.

Insert Table 8 here

So in theory the total ratings sampling errors are independent of the size of the ratings which make up the schedule. Certainly the variations in the percentage sampling errors are nothing like the variations in the single spot percentage sampling errors. In practice the equality of the schedule total ratings sampling errors will depend upon correlations in viewing between spots.

We can get a feel for whether or not this works in practice by comparing the ITV and Channel 4 components of each schedule from section 2.3. On average, single ITV ratings are about three times as high as single Channel 4 ratings. In table 9 the schedule components have been ranked according to total ratings delivered.

Insert Table 9 here

Although the evidence is not exactly in line with the hypothesis that schedules with equal total ratings have equal sampling errors, it is certainly not the case that a schedule of low rating/high percentage sampling error spots will have correspondingly high sampling error for the total ratings.

But what about a restricted availability channel?

To generate the same impacts, a restricted availability channel with a 20% penetration would need to generate ratings five times as large within its own universe, ie. 1,000 ratings in total.

The equivalent schedule structures and theoretical sampling errors, now based upon a sample of only 92 Men, are shown in table 10.

Insert Table 10 here

Even with the ridiculously high 50 rating spots, the order of magnitude of the sampling errors is preserved.

This time the empirical evidence shown in table 11 is very thin, with only two satellite schedules coming close to the terrestrial channels' total ratings levels.

Insert Table 11 here

However, these seem to fit in with the hypothesis of equal sampling errors for equal schedule total ratings.

The last hypothesis considered is: Do schedules with equal impacts have the same sampling error for total ratings irrespective of the sample size of the demographic sub-group analysed? This is analogous to the restricted availability channel situation.

Again we can get a feel for whether or not this works in practice by re-examining the schedule sampling errors shown in section 2.3. This time the Men 16-24 total ratings are multiplied by 62/459 (the ratio of the sample sizes) to form percentages of the All Men universe (equivalent to a comparison of schedule total impacts) before ranking according to the total ratings delivered, as shown in table 12.

Insert Table 12 here

There are some exceptions but in general the hypothesis holds. More importantly, it is certainly not the case that Men 16-24 sampling errors are 2.7 times as big $\left(=\sqrt{459/62}\right)$ as equal impact All Men sampling errors.

4.3 Summary

Many applications of TV peoplemeter panel data result in the need to measure small audiences. These may involve regional or demographic sub-group analysis as well as low rating or restricted availability channels. In these situations it is important to understand the sampling errors involved so that the best use of existing panel data is made.

The sampling errors associated with audience measurements of individual minutes or commercial spots are often huge. In the context of advertising schedules, any attempt at optimising the choice of individual spots is often unjustified.

However, it is well known that the total ratings for whole schedules have much lower sampling errors than the individual spots within a schedule. In fact it is broadly true that schedules with equal impacts have equal sampling errors irrespective of the size of the individual spot ratings or the sample size of the sub-groups analysed. Of course there are limiting situations in which this equality breaks down, but these would correspond to unusually heavy advertising on an individual channel.

Undoubtedly this finding will be useful in many situations. However, it cannot be allowed to generate complacency. In practice, a schedule on a low rating, restricted availability channel would never generate total impacts for a sub-group which were equal to those for a main category on a high rating, national channel.

Within the existing panels, there are measurements which provide a significantly more robust basis on which to trade advertising airtime. However, in many cases there is still no substitute for increased sample size.

5. Alternative Measurement Methods for Low Penetration Channels

5.1 Choices

In a recent paper (Franz 1997) points out that small rating stations may get neglected in media planning because of their low representation on peoplemeter panels. He suggests using independent samples collecting data say monthly and capable of being aggregated to large numbers of respondents in a year.

One advantage is that a large sample size made up of independent samples reduces bias which may be significant for small stations on a permanent panel. Out-of-home viewing may also be included in the measurement. He advocates normalising the viewing levels to panel levels so that data can be used comparably (presumably separating out-of-home viewing). The use of data aggregated over time would mean that the data would be for strategic planning and the panel data would provide some tactical information.

The techniques listed by Franz for a strategic television monitor are:

- personal interviews: paper or CAPI
- computer aided telephone interviews
- self-completion diaries

We have a case history to report from the UK using self-completion diaries, not in the continuous way suggested by Franz but as periodic snapshots.

5.2 Broadband Cable Audience Diary Research: A Case History

Prior to November 1997, the BARB TV peoplemeter panel operation in the UK did not publish any audience estimates based upon the broadband cable universe because the sample sizes available were considered too small. Even now, only the five larger cable exclusive channels are itemised in the reporting system. Therefore, since January 1996, the UK's Cable Research Group have commissioned RSMB to conduct periodic diary based studies to fulfil the broadband cable industry's audience measurement requirements. So far, four such studies have been completed (January 1996, September 1996, March 1997 and September 1997) and two more are due in 1998 (March and September).

It must be acknowledged that a paper diary is an inferior data collection mechanism when compared to a peoplemeter. However, the counter-argument is: what is the point of having a very precise measurement of small audiences in only a very small sample peoplemeter panel with consequently huge sampling errors? Because it is cheaper and larger samples are therefore more affordable, a diary based study can be a more cost effective solution. Significantly the Cable TV (CATV) audience research has been formally approved by the Institute of Practitioners in Advertising (IPA) and generally accepted as providing a valid audience measurement.

The latest study (September 1997) was based upon a sample of 1300 adults and 400 children, each completing a two-week quarter hour viewing diary covering all channels available in broadband cable homes. Whilst this sample size is effectively double that available from the BARB panel, we should not pretend that this completely solves the small audience measurement problem. The sampling errors for individual channels are still large. However, the value of the CATV survey is not all about increased sample size:

- The diary sample is selected from an Establishment Survey of 2,500 households. This provides up-to-date information about the penetration and demographic profile of each broadband cable channel.
- Identification of the cable channels received by each diary respondent allows analysis of viewing behaviour within receivers of each channel. This is not possible within the peoplemeter panel operation.

- Following the diary recording period, each respondent completes a leave behind questionnaire. This is designed to collect information on usage of other media, opinions of individual channels and impressions of cable operator services, information which could not be collected from an audience measurement panel. In this way, the service is tailored to the needs of the members of the Cable Research Group.

- By boosting the sample it is possible to provide audience measurement data for individual cable operator areas.

A potential disadvantage of the short term diary study is its inability to measure schedule reach beyond two weeks. This is overcome with the usual probability modelling techniques which are commonly employed within radio and press research.

6. Factoring Regional Sub-Group Audiences

An approach under test in the UK for the estimation of regional sub-group audiences is that of factoring from the Network panel.

The basic principle is that for a particular minute, the conversion from a main category to a sub-group rating is the same within any region as it is within the larger network panel. In algebraic terms:

$$\frac{s}{m} = \frac{S}{M}$$

where:

s	= sub-group rating, small area.
m	= main category rating, small area.
S	= sub-group rating, network.
M	= main category rating, network

By re-arranging the above formula, we can derive the basic factoring model for estimating sub-group ratings in small sample areas:

$$s = \frac{S}{M} \times m$$

The model is improved by weighting the network panel to the small area panel profile before calculating the network ratings S and M.

This approach inevitably leads to results which are more stable than actual small sample based data, because all the components on the right hand side of the formula are based upon larger samples. However, if the principle is not valid, then results will be biased. The purpose of the evaluation is to examine the trade-off between improved stability and potential bias.

6.1 Theoretical Reductions in Variability (Sampling Error)

The sampling error of a factored rating will be a function of the sampling errors of all three components of the factoring formula and their correlations. The mathematics of the theory are quite tortuous, but if we make some fairly well justified assumptions then the following relatively simple formula can be derived for the relationship between the sampling errors for factored and actual ratings:

$$\frac{\text{Factored sub-group s.e.}}{\text{Actual sub-group s.e.}} = \sqrt{\frac{n_s}{n_m} + \frac{n_s}{N_s} + \frac{n_s}{N_m}\left(1 - \frac{2\bar{x}_s}{\bar{x}_m}\right)}$$

where:

n_s	= small area sub-group sample size
n_m	= small area main category sample size
N_s	= network sub-group sample size
N_m	= network main category sample size

$\bar{x}_s$ = expected sub-group rating
$\bar{x}_m$ = expected main category rating

The problem with this formula is that it depends upon the expected ratings levels. However, in the UK example where n_s (small area sub-group sample size) is small compared with N_m (network main category sample size), an adequate approximation is:

$$\frac{\text{Factored sub-group s.e.}}{\text{Actual sub-group s.e.}} = \sqrt{\frac{n_s}{n_m} + \frac{n_s}{N_s}}$$

As an example, consider Men 16-24 in London. The associated main category is All Men. First we need the relevant sample sizes. The following are effective sample sizes which reflect the weighting required to correct for geographical disproportionate sampling, demographic disproportionate sampling and other "accidental" panel imbalances:

n_s = London Men 16-24 effective sample size = 49
n_m = London All Men effective sample size = 322
N_s = Network Men 16-24 effective sample size = 421

Then:

$$\frac{\text{Factored sub-group s.e.}}{\text{Actual sub-group s.e.}} = \sqrt{\frac{49}{322} + \frac{49}{421}} = 0.52$$

In this example the sampling error is nearly halved - there is a theoretical reduction in variability of 50%. This is equivalent to a four-fold increase in the panel sample size for this sub-group.

For larger sub-groups, the reductions in sampling error are obviously not so great, as shown in table 13.

Insert Table 13 here

6.2 Practical Reductions in Variability - Analysis of Variance

For applications of the TV audience measurement data which involve comparisons of sub-group ratings between regions, the sampling error approach to the assessment of variability reduction is appropriate. However, for applications which involve change over time within a single region, it must be noted that a component of the sampling error comes from the initial selection of the sample. When this sample is used on a continuous basis as a panel, the initial recruitment sampling error is equivalent to an ongoing 'bias'. The resulting sampling errors on measurements of change over time are consequently smaller.

The analysis of variance procedure is designed to generate a practical rather than a theoretical measurement of the reduction in variability achieved through factoring. Published and factored ratings are calculated for every quarter hour, for every channel, for every day, for every week and input into the analysis of variance procedure. After allowing for as many known "systematic" variations as possible (e.g. the daily quarter ratings pattern, the differences between channels) and their interactions, the analysis of variance procedure calculates a residual variance. This is taken to be the average variability for any particular quarter hour measurement and is used to compute the associated coefficient of variation for a typical quarter hour rating. This is analogous to the percentage sampling error for a quarter hour rating. Then we can calculate the percentage reduction in this coefficient of variation for factored ratings against published ratings. Example results for London and shown in Table 14.

Insert Table 14 here

So in terms of change over time, the reductions in variability are still worthwhile if not so large. The full benefits of factoring will only be realised in comparisons between regions when the initial recruitment sampling error component is also relevant.

6.3 Potential Bias

In practice it is very difficult to determine whether or not factored results are biased. The prediction error will be a mixture of model bias and random error which are difficult to untangle due to the large sampling errors associated with the actual sub-group measurements. All we can do is compare factored and actual results at various levels of detail and to search for exceptional differences. If exceptional differences are always at times when the regional programming is different to the network, then there may be a problem. Otherwise we have to judge the relative credibility of factored and actual results. Remember that in many cases factoring is designed to replace "unbelievable" and erroneous results with more credible audience measurements - by definition these would be different.

At the highest level of aggregation, table 15 compares factored and actual four week all-time average ratings for Total TV and ITV in London in March 1996.

Insert Table 15 here

All these differences between factored and actual ratings are within sampling error. The largest differences are for Women AB. However, at this time the actual data showed that Women AB viewing in London was 20% lower than in the whole network. Although viewing levels in London are expected to be lower, the factored data seems to provide a more credible result.

At greater levels of detail, the differences factored and actual ratings are obviously greater but still within sampling error. Another way to evaluate the factoring model is by examining exceptional differences at the quarter hour level.

For example, amongst Men 16-24 in London, the biggest difference between factored and actual quarter hour ratings was on BBC1 at 7:45pm on Thursday 21st March. At this time the actual rating was 13% and the factored rating was 26%. The first thing to note is that the same programme was being shown in London and across the whole network - this is not a bias caused by inconsistent programming. To put this exceptional difference into context, table 16 shows the actual and factored ratings in adjacent weeks:

Insert Table 16 here

In this case, the factored rating provides a more credible result in relation to the adjacent weeks, a finding repeated for all the exceptions examined so far. However, it should be noted that our examination of exceptions has been based upon factoring from a reduced network panel which may minimise programme schedule effects.

6.4 Summary

The factoring approach to small area sub-group audience measurement is still under test in the UK. The advantages are significant in terms of reduced variability because factoring is equivalent to adding between 50% and 100% to the current panel sample sizes but at virtually no additional cost.

Across a wide range of sub-groups and "difficult" areas, we have so far found no evidence of bias in the factoring model. Analysis of exceptions always points to more credible factored results and factoring is no worse during times of inconsistent programming between a region and the network. The potential disadvantages are that unforeseen changes in regional programming policy could disrupt the factoring principle and that unfactored sub-group data would always be available to support any criticism of factored results.

The issue of potential bias is still under investigation and if the results are positive, then factoring could provide a real solution to the small area sub-group audience measurement problem.

7. Summary and Conclusions

1) Mass audience commercial television channels created the need for metered panels, ultimately using peoplemeters almost universally.

   Peoplemeters measure audience movements very precisely in terms of time, necessary for advertisement audiences but at the expense of sample size.

   Trading in television has been initially centred around the audiences to single spots.

2) Sample size has become a severe limitation, in relation to the way in which the medium has developed, with requirements to measure ever smaller audiences. The most extreme problems are for advertisers but occur for the programming side as well.

3) This paper distinguishes between different ways in which small audiences occur, considers the reliability of the data from existing systems (ie. sampling error) and suggests possible strategies for dealing with the problem.

   Small audiences occur increasingly within mass audience research systems for:

   - large stations at off peak times and for small sub-groups.
   - smaller stations operating within the same universe.

4) A great deal of effort and money is wasted pursuing the single spot philosophy in situations where data are unreliable. The problems occur much higher up the scale of audience size than is always realised. The paper gives examples of sampling errors to demonstrate this.

5) The solutions we suggest are:

   - it is worth considering whether the extra cost of increasing sample sizes might in fact save money.
   - planning, trading and accountability need to move away from single spots to aggregated data either over weeks for single times or across whole schedules.
   - the use of factoring for sub-group data is examined for the special case of a regionally fragmented network system.

6) Digital television will create an initially small but important growing universe. Undoubtedly this will require, initially separate, boosted peoplemeter panels.

   The number and complexity of channel choice, however, will create audience fragmentation.

   Peoplemeter data will have to be used in an entirely different way. This will involve setting limits to the reach and share of channels. Most advertising will need to be assessed on aggregated data and probably schedules, extending across a range of channels. Variations in spot audiences will be meaningful in only a small number of cases. Some programming data will be usable if aggregated over time.

7) Channels available to other smaller universes where meters are unaffordable, in some cases, can use alternative techniques such as paper diary panels or recall. These are ideal for evolving markets where a snapshot at a point in time provides the clearest understanding of a potential audience’s response to a channel.

8) For markets which have reached a relatively stable position, surveys accumulating large samples over time may provide the basis for strategic decisions, potentially providing more reliable representation of minority interests. The problem will be the growing complexity of choice and the use of electronic programme guides which may undermine the element of awareness and recall involved in non-meter techniques.

9) We believe that a crucial principle has been put forward in this paper.
Sampling error is fairly similar for the same number of rating points no matter whether arising from a single spot or a schedule.

Measurement of small audiences can therefore become as reliable as for large audiences when the small audiences are combined together.

10) So for the fragmented audiences of the future, research systems have to change and also the ways in which the research is used. Research users cannot go on looking at smaller audience forms by turning up the magnification of a limited microscope and seeing ever more blurred pictures.

References:

American Research Bureau Inc. New York, “Arbitron replication: a study of the reliability of broadcast ratings”, 1974.

Boon, A.K. den “The reliability of television audience ratings”., in ARF/ESOMAR Worldwide Electronic and Broadcast Audience Research Symposium, 1994, Paris, France.

Franz, G. “How to catch small fish approaches to the measurement of small reach stations”, in: ASI 1997 European Television Symposium, 1997, Budapest, Hungary.

Read, S. & Johnson, J. “Audience measurement in the 21st Century”, in: ASI 1997 European Television Symposium, 1997, Budapest, Hungary

Schillmoeller, E.A. “Audience estimates and stability”, in: ARF/ESOMAR Worldwide Broadcast Audience Research Symposium, 1992, Toronto, Canada.

Twyman, T. & Wilcox, S. “The variability of audience measurement data and how to live with it”, in: ARF/ESOMAR Worldwide Electronic and Broadcast Audience Research Symposium, 1996, San Francisco, USA.

Wilcox S. & Reeve B. “Statistical efficiencies in the new UK television audience measurement panels”, in: ARF/ESOMAR Worldwide Broadcast Audience Research Symposium, 1992, Toronto, Canada.

The Authors:

Tony Twyman is Technical Advisor to BARB: Broadcaster’s Audience Research Board and AIRC: Association of Independent Radio Companies, UK

Steve Wilcox is Technical Director at RSMB Television Research Limited, UK

Table 1 Sample Sizes and Sampling Error
London, ITV Single Minute Rating, Monday, 8:30pm

	Sample Size	TVR	95% Confidence Interval
All Individuals	1204	9.6	±24%
Adults	996	11.1	±22%
Men	459	11.0	±28%
Women	537	11.2	±26%
Housewives	530	12.7	±22%
Housewives with Children	154	9.8	±49%
Women ABC1	296	8.9	±42%
Men 16-34	164	5.9	±70%
Women AB	129	4.8	±88%
Men 16-24	62	7.8	±88%
Children	208	2.5	±105%

Table 2 Channel Share Sampling Errors - All Adults 16+, Network

		95% Confidence Interval		
Channel	**Share**	**Single Minute**	**Single Day**	**4 Week Average**
BBC1	30%	±10%	±1.4%	±0.9%
BBC2	11%	±19%	±2.3%	±1.2%
ITV	33%	±9%	±1.2%	±0.8%
Channel 4	10%	±20%	±2.1%	±1.2%
Channel 5*	3%	±30%	±5%	±3%
Satellite	1-1½%	±50%	±8%	±5%
Satellite	½-1%	±65%	±10%	±6%
Satellite*	Under ½%	±90%	±15%	±8%
S4C*	0.3%	±90%	±15%	±8%
Cable only*	Under ¼%	±100%	±20%	±10%

Table 3 Channel Share Sampling Errors - 10% Sub-Group or Region

		95% Confidence Interval		
Channel	**Share**	**Single Minute**	**Single Day**	**4 Week Average**
BBC1	3.0%	±32%	±4%	±3%
BBC2	1.1%	±61%	±7%	±4%
ITV	3.3%	±29%	±4%	±3%
Channel 4	1.0%	±64%	±7%	±4%
Channel 5*	0.3%	±96%	±16%	±9%
Satellite	0.10-0.15%	±160%	±25%	±16%
Satellite	0.05-0.10%	±208%	±32%	±19%
Satellite*	Under 0.05%	±288%	±47%	±25%
Cable only*	Under .025%	±320%	±63%	±32%

Table 4 Channel Share Sampling Errors - 10% Sub-Group in a 10% Region

		95% Confidence Interval		
Channel	**Share**	**Single Minute**	**Single Day**	**4 Week Average**
BBC1	0.3%	±100%	±14%	±9%
BBC2	0.1%	±190%	±23%	±12%
ITV	0.3%	± 90%	±12%	±8%
Channel 4	0.1%	±200%	±21%	±12%

Table 5 Sampling Errors for Individuals Minute Ratings

Channel	Time	All Men		Men 16-24	
		TVR	95% c.i.	TVR	95% c.i.
ITV	7:45am	1.3	±39%	2.5	±95%
	1:45pm	5.1	±21%	3.8	±68%
	8:30pm	11.0	±14%	7.8	±44%
CH4	7:45am	0.5	±81%	0.0	-
	1:45pm	0.4	±77%	0.0	-
	8:30pm	5.0	±23%	0.0	-

Table 6 Sampling Errors for Average Ratings - Four Mondays

Channel	Time	All Men		Men 16-24	
		TVR	95% c.i.	TVR	95% c.i.
ITV	7:45am	1.2	±24%	1.6	±57%
	1:45pm	3.5	±15%	2.1	±45%
	8:30pm	13.2	± 5%	7.1	±19%
CH4	7:45am	0.4	±39%	0.1	±88%
	1:45pm	1.0	±13%	0.4	±43%
	8:30pm	3.1	±10%	1.3	±33%

Table 7 Sampling Errors for Schedule Total Ratings

Schedule	Channel	All Men		Men 16-24	
		Total TVRs	95% c.i.	Total TVRs	95% c.i.
I	ITV	33	±15%	39	±25%
	Channel 4	19	±18%	15	±51%
	Satellite	4	±26%	2	±67%
	Total	55	±12%	56	±29%
II	ITV	148	±6%	87	±19%
	Channel 4	71	±9%	54	±24%
	Satellite	14	±18%	16	±52%
	Total	233	±5%	157	±16%
III	ITV	140	±7%	116	±18%
	Channel 4	36	±9%	28	±24%
	Satellite	10	±19%	28	±33%
	Total	186	±6%	173	±15%
IV	ITV	254	±6%	134	±20%
	Channel 4	171	±6%	148	±21%
	Satellite	36	±17%	55	±28%
	Total	461	±5%	337	±15%
V	ITV	174	±7%	88	±26%
	Channel 4	64	±9%	32	±29%
	Satellite	17	±20%	21	±58%
	Total	256	±6%	141	±21%

Table 8 Sampling Errors for Schedules with Different Structures

Single Spot TVR	Single Spot % s.e.	Number of Spots	Total TVRS	Total TVRs 95% c.i.
20	2.2%	10	200	±5.9%
10	3.2%	20	200	±6.3%
5	4.7%	40	200	±6.4%
1	10.7%	200	200	±6.6%
½	15.2%	400	200	±6.6%

Table 9 Sampling Errors for ITV and Channel 4 Schedules

Schedule	Channel	All Men	
		Total TVRs	95% c.i.
IV	ITV	254	±6%
V	ITV	174	±7%
IV	CH4	171	±6%
II	ITV	148	±6%
III	ITV	140	±7%
II	CH4	71	±9%
V	CH4	64	±9%
III	CH4	36	±9%
I	ITV	33	±15%
I	CH4	19	±18%

Table 10 Sampling Errors for Schedules on Restricted Availability Channels

Single Spot TVR	Single Spot % s.e.	Number of Spots	Total TVRS	Total TVRs % s.e.
100	-	10	1000	-
50	2.4%	20	1000	±4.7%
25	4.2%	40	1000	±5.7%
5	10.5%	200	1000	±6.4%
2½	15.1%	400	1000	±6.5%

Table 11 Sampling Errors for Satellite Schedules

Schedule	Channel	All Men	
		Total TVRs	95% c.i.
IV	Satellite	36	±17%
V	Satellite	17	±20%

Table 12 Sampling Errors for Main Categories vs. Sub-Groups

Schedule	Channel	Category	Total TVRs	95% c.i.
I	Total	All Men	55	±9%
IV	**Total**	**Men 16-24**	**46**	**±15%**
III	Channel 4	All Men	36	±9%
IV	Satellite	All Men	36	±17%
I	ITV	All Men	33	±15%
III	**Total**	**Men 16-24**	**23**	**±15%**
II	**Total**	**Men 16-24**	**21**	**±16%**
IV	**Channel 4**	**Men 16-24**	**20**	**±21%**
I	Channel 4	All Men	19	±18%
V	**Total**	**Men 16-24**	**19**	**±21%**
IV	**ITV**	**Men 16-24**	**18**	**±20%**
V	Satellite	All Men	17	±20%
III	**ITV**	**Men 16-24**	**16**	**±18%**
II	Satellite	All Men	14	±18%
II	**ITV**	**Men 16-24**	**12**	**±19%**
V	**ITV**	**Men 16-24**	**12**	**±26%**
III	Satellite	All Men	10	±19%
I	**Total**	**Men 16-24**	**8**	**±29%**
II	**Channel 4**	**Men 16-24**	**7**	**±24%**
IV	**Satellite**	**Men 16-24**	**7**	**±28%**
I	**ITV**	**Men 16-24**	**5**	**±25%**
I	Satellite	All Men	4	±26%

Table 13 Theoretical Reductions in Variability - London

	Penetration of Main Category	Reduction in Sampling Error	Equivalent Sample Increase
Housewives with Children	29%	27%	x1.9
Women ABC1	55%	8%	x1.2
Men 16-34	36%	28%	x1.9
Women AB	24%	28%	x1.9
Men 16-24	14%	48%	x3.7

Table 14 Practical Reductions in Variability - London

	Penetration of Main Category	Reduction in Sampling Error	Equivalent Sample Increase
Housewives with Children	29%	17%	x1.5
Women ABC1	55%	6%	x1.1
Men 16-34	36%	14%	x1.4
Women AB	24%	25%	x1.8
Men 16-24	14%	35%	x2.4

Table 15 Four Week Average Ratings - Factored vs. Actual - London

	Total TV	ITV
Housewives with Children	+1%	+1%
Women ABC1	+1%	+3%
Men 16-34	-3%	-4%
Women AB	+7%	+8%
Men 16-24	-2%	-2%

Table 16 Exceptional Difference - London - Men 16-24
Thursday 7:45pm - 8:00pm

	Actual Rating	Factored Rating
Week 1	21	25
Week 2	21	30
Week 3	**13**	**26**
Week 4	21	27