

CONTENTS

I	SUMMARY	1
II	BACKGROUND	4
III	OBJECTIVES	5
IV	TEN PRINCIPLES	7
V	ORGANIZATION PRINCIPLES, CONTROL AND FUNDING	9
VI	DATA COLLECTION	11
1.	The Universe	12
2.	The Issue of Quality	13
3.	Towards a High Quality Sample: Probability Sampling	13
4.	The Sampling Frame.....	14
5.	The Multi-Stage Sample Design – and Stratification	15
6.	Establishment Surveys	16
7.	Panel Recruitment, Maximizing Response, Response Rates	19
8.	Panel Controls.....	23
9.	Sample Size	25
10.	Panel Incentives.....	26
11.	Panel Maintenance and Renewal.....	26
12.	TV Audiences to be Measured.....	28
13.	The Definition of “Audience”	28
14.	The Peoplemeter Technology	30
15.	Guest Viewing	30
16.	Nil Viewing – and Panelists on Holiday	31
17.	The Diary.....	32
18.	Quality Control	35
19.	Monitoring System Performance.....	37
20.	Coincidental Studies	39
21.	Editing Rules	40
22.	Weighting.....	41
23.	The Independent Audit	42
24.	Additional Data Collection – and Program Appreciation	42
25.	Viewing outside the Home	43
26.	International and Niche Channels	44
VII	DATA REPORTING	45
1.	Universe Estimates.....	45
2.	A Viewer: Calculation Convention	47
3.	Components of Viewing	49
4.	Audience Measures	50
5.	Availability of Data for Analysis	55
6.	Speed of Access and Reporting.....	57
7.	Reporting of Channels	58
8.	Universes for Reporting	58
9.	Program Classification.....	59
10.	Demographic Variables for Analysis	59
11.	Small Samples.....	61
12.	Time Units for Analysis	61

VIII THE FUTURE..... 63

ANNEX: THE AUDIENCE RESEARCH METHODS GROUP..... 66

I SUMMARY

1. Every country or market needs **valid and reliable data** on television audiences. By **valid** we mean that the measurement system employed to measure television audiences actually *measures what it purports to measure*. By **reliable** we mean that the measurement system would *yield very similar findings if independently carried out a number of times*. Senior management and schedulers within broadcasting organizations require accurate information on their own audiences and those of their competitors – as an input to program content and scheduling decisions. For channels funded wholly or partly by advertising, TV Ratings are the “currency” for the buying and selling of commercial airtime. For channels funded in whole or in part by a public licence fee or the equivalent, audience data provide measures of public accountability. Subscription channels equally need audience data. With respect to channels for which ratings data are the currency for commercial airtime, these guidelines represent the latest in a series of efforts to obtain an international “exchange rate” for TV audience measurement systems throughout the world. **As such, they are considered part of a process that will lead to further revisions and improvements – reflecting changes in TV technology, research technology and user needs of TV audience measurement systems.**
2. The technology of broadcasting and the trading of goods and services are increasingly international in nature. The **international** advertiser needs to be able to compare audiences across national boundaries. International comparisons form the basis for advertising investment decisions between markets. There are major issues of **harmonization** to be tackled, to encourage as much comparability as possible in TV audience measurement systems between and among markets.
3. The key issue is the need for quality and a degree of standardization in the **data collection** procedures of TV audience measurement systems. This is the critical route to enable between-country comparisons to be made. A balance must be struck between standardization across measurement systems and a development path to allow migration to improved measurement approaches which will be required for the future.
4. For **reporting** of TV audiences, individual countries will want to retain certain traditional local conventions, for example the appropriate definition of “prime time” for their market. The key harmonization issue for data reporting is then the availability **in addition** of audience data that conform to international reporting conventions – to enable users to make between-market comparisons.

5. A number of techniques exist to measure TV audiences – from the comparatively simple to the highly complex. The “**peoplemeter**” system has become the de facto standard audience measurement system throughout the more developed markets around the world. But other techniques also have their place: the self-completion diary, recall, the coincidental interview – even observation.
6. It is the purpose of these guidelines, drawn up by an international committee of experts known as the **Audience Research Methods (ARM) Group** and published by the EBU, to provide **operational recommendations** on how such techniques should be designed.
7. The emphasis, however, is upon the **peoplemeter** system. This is the one in which nearly all developed markets have invested most heavily. It is a sophisticated system able to deliver a wide range of audience measures. Equally it meets a wide range of management needs – both editorial and commercial. It is however a complex and relatively expensive system to operate. It involves the equipping of a probability sample of TV households with:
 - Meters to establish when each TV set and (usually) VCR is switched on and to which channel and program it is tuned.
 - Handsets whereby all individuals (usually down to the age of 4) register when they personally are watching television.
8. The operational guidelines are based on ten over-riding **Principles**:
 - Meeting total marketplace needs
 - Effective industry consultation
 - Full disclosure
 - Optimal resource allocation
 - Scientific method
 - Best research practices
 - Quality control
 - Maximising response
 - Equal access to data
 - Methodological experimentation

All users of the service need to have **confidence** in the system. This requires full consultation, to ensure that the best technical solutions are adopted for their own market. These must be in the public domain. In the interest of fair trading all user groups should share the same conditions of open access to the data.

9. How a country’s audience measurement system is **designed, controlled and funded** is another critical issue. **The key principle is that there needs to be effective consultation with all sectors of the industry, to ensure that the system is customer-led in its priorities.** Each country needs to determine what

precise organizational arrangement legally and best fulfils this objective – a continuum ranging from: the joint industry committee (JIC) that inherently ensures that the interests of all users of the system are represented in its management, to a media owner committee (MOC), to a particular data supplier’s own service (OS).

10. The substance of these guidelines is, however, the **operational detail** recommended for:

- Sample design
- Sample recruitment and maintenance
- Data collection
- Data reporting
- Quality control
- Transparency of methods used

11. Recognising the need for updating, the guidelines conclude with pointers to the **future**: in particular how TV audience measurement will need to respond to the challenges of an increasingly fragmented and complex TV market, and to the new technologies such as digital and cyberspace.

12. The ARM Group, who drew up these guidelines representing all sectors of the industry:

- Television broadcasters
- Advertisers
- Advertising agencies

commend them to the industry as “**Towards Global Guidelines for Television Audience Measurement**”.

II **BACKGROUND**

These guidelines are the third edition of the successful and widely distributed “Green Book” entitled: “Towards Harmonization of Television Audience Measurement Systems”. The second edition was published in 1993 under the joint auspices of:

- European Broadcasting Union (EBU) – as sponsor and publisher
- Association of Commercial Television in Europe (ACT)
- European Association of Advertising Agencies (EAAA)
- European Group of Television Advertising (EGTA)
- European Organizations for Media Research (EMRO)
- Group of European Audience Researchers (GEAR)
- World Federation of Advertisers (WFA)

While global in context and involving the WFA, it was essentially conceived and created within Europe.

This edition has been prepared with the active support and involvement of the above key industry organizations – with the important addition of:

- USA’s Advertising Research Foundation (ARF)
- Canadian Advertising Research Foundation (CARF)
- European Society for Opinion & Marketing Research (ESOMAR)
- Pan-European Television Research Group (PETV)

As with earlier editions, involvement has also been sought at the drafting stage from the principal data suppliers around the world. A wide range of valuable comments and suggestions have been received from them and these have been taken into account or incorporated in the text of this third edition.

It breaks new ground in incorporating the perspectives and experiences gained both in the Americas and in Australasia and the developing countries of the world. This global perspective is reflected in the name for the new edition of these guidelines:

“Towards Global Guidelines for Television Audience Measurement”.

This third edition is part of a process that will lead to further revisions and improvements – reflecting changes in TV technology, research technology and user needs of these systems.

III OBJECTIVES

In practically all countries in the world television has long been, or is fast becoming, a dominant medium for delivery of information, entertainment and commercial communication – if not the principal one. All parties to television:

- TV broadcasters and program producers
- TV company sales houses
- Advertisers
- Advertising agencies and media specialists
- Government
- Equipment manufacturers
- The academic community
- The TV viewing public

want valid and reliable information that accurately and consistently reflects the performance of individual channels and programs, in terms of the size and characteristics of their audiences. This is true whether the revenue source for the channel is advertising, subscription, pay per view or some form of licence fee or equivalent public funding.

In parallel, trans-frontier marketing has become a part of contemporary business life. Both television companies and the multi-national advertising community are calling for standardized audience data to make international comparisons and analyses possible. They see a need for standard methodological procedures and user conventions to improve the comparability of national or individual market audience data, facilitating the free flow of audience information across borders.

This document has, therefore, been designed to answer the many needs of a wide variety of user groups. Directed towards the whole of the international community, it endeavours to recognise both similarities and differences which exist across national boundaries. However, underlying these guidelines is a belief that the needs and aspirations of the users do not vary by geography.

The following types of organization are explicitly identified as potential beneficiaries of these guidelines:

- Terrestrial, cable and satellite TV networks
- Operators of cable and satellite delivery systems
- TV stations and video distributors
- TV program production companies
- Television advertisers
- Advertising agencies
- Research companies

The document should be found useful at a number of levels in such organizations:

- Senior corporate management
- Marketing and sales management
- Practising professional researchers

Accordingly, the overall objectives of these guidelines are to:

- Establish an international consensus on the research methods to be used to deliver the most valid and reliable audience estimates for the television medium. To reiterate, by **valid** we mean that the measurement system employed to measure television audiences actually ***measures what it purports to measure***. By **reliable** we mean that the measurement system would ***yield very similar findings if independently carried out a number of times***.
- Identify and publish good professional practice for the design and operational procedures of TV audience measurement.
- Identify and where possible discourage practices that fall short of proven standards where well-founded evidence exists, while accepting different practices where conclusive evidence of the superiority of one particular approach is not available.
- Encourage a commitment by all sectors of the TV research community to such standards as will enable all users of the medium to access and compare data across the globe on a comparable basis.
- Create a continuing international research forum to stimulate improved methodologies for the collection and reporting of TV audiences, in the face of constantly changing world economies, TV markets and video technologies.

IV TEN PRINCIPLES

Underlying the framework and the detail of the operational guidelines are ten basic Principles that apply generically to media research and specifically to TV audience measurement:

1. Meeting Total Marketplace Needs

The television audience measurement system needs to be designed to cover the total TV marketplace of the country or market in question. It should set out to meet the needs of all categories of users of such information.

2. Effective Industry Consultation

Full consultation should at all times take place between the research company/ies and all users of TV audience information. In countries where joint industry bodies manage the system, consultation is inherent to the structure. In other markets specific steps are required to ensure that consultation with the user community is systematic and effective. Whatever the organizational structure in a particular market, joint industry technical advisory bodies can and should exist, to ensure an ongoing collective dialogue with clients.

3. Full Disclosure

The full detail of methodological procedures shall be openly available to all subscribers. The user is then able to understand and make allowances for whatever compromises have been necessary in a particular market, where the system falls short of the ideal.

4. Optimal Resource Allocation

Research resources should be deployed effectively to take account of the balance of requirements of the commercial and editorial management decisions to be based upon the information provided.

5. Scientific Method

The research methods need to be tried and tested, and scientifically based. It is important to strive for system **validity** and **reliability**. By system validity we mean that it ***actually measures what it purports to measure***. By reliability we mean that it would ***yield very similar findings if independently carried out a number of times***.

6. Best Research Practices

For most elements of data collection and reporting there is an ideal procedure that should be observed where possible. While departures from this ideal usually involve compromises, there may be other acceptable procedures that can and should be adopted. In all circumstances the principle of Full Disclosure (above) needs to be rigorously observed.

7. Quality Control

Rigorous and systematic quality control procedures need to be deployed for each element of fieldwork, data collection, editing and reporting. Audience measurement systems would normally be expected to conform to all relevant international codes of conduct (e.g. for Europe, the ESOMAR Code of Conduct).

8. Maximising Response

The burden placed on respondents should be minimised in the interest of high response rates, minimum exposure to bias and towards the gathering of valid and reliable information. Equally, the respondent's right to privacy and confidentiality shall be respected at all times.

9. Equal Access

It is in the interest of fair trading that all user groups share the same conditions and a fair price for access to audience data, contributing thereby to:

- Openness of the measurement systems themselves.
- Even trading conditions between buyers and sellers of TV airtime.
- Maximum exploitation of the data that have been relatively costly to collect.

10. Methodological Experimentation

Research organizations are encouraged to be innovative, and in particular to conduct carefully controlled experiments of alternative procedures. In particular they need to conduct regular systematic investigations of the implications of non-response and response error to their systems. Their methodologies and outcomes shall be fully documented and publicly available to all user groups.

V ORGANIZATION PRINCIPLES, CONTROL AND FUNDING

The supervisory organization of audience measurement has an important part to play in ensuring the production of data that meet the requirements of all users. In principle three types of organizational structure exist – though in practice around the world there are variations around each, leading to a continuum of organizational arrangements.

Research Company's Own Service (OS). A research company supplies audience data as a private commercial venture, and signs multiple individual contracts with purchasers of the data. It is expected that OS Systems would make formal provision for regular user consultation via user groups, and be subject to independent audit and accreditation procedures involving active participation of the industry.

User Sector Committee. A number of users award a contract for a specific service and thereby guarantee the funding. A common example of this type of structure is the collective of broadcasting organizations: **Media Owner Committee (MOC).** Advertisers and agencies are then not involved in specifying the terms of the MOC licence nor in supervising the contract, but they should be invited to participate in consultative user technical committees. Failure to engage in formal and regular consultation with all users of the system, in particular advertisers and agencies, would be unacceptable. The media owners may hold the copyright, or may permit the research company to retain the copyright and sell data to other parties.

Joint Industry Committee (JIC). The research company/ies conducting the fieldwork and data processing hold(s) a contract with a formal Joint Industry Committee of representatives from the broadcasters, the advertisers and the agencies. The JIC typically draws up a specification for the service, invites tenders, awards the contract, supervises the service, owns the copyright and determines the licensing conditions and terms of access. The day-to-day management and technical functions will often be delegated to representative management and technical advisory committees.

Provided all sectors of the industry are represented, the principle of Effective Industry Consultation is inherent to a JIC structure. It ensures that all sectors of the industry are consulted and involved in the design and management of the system. This is especially important because in any particular market funds will rarely be available to support more than one audience measurement system. It ensures that the research system is customer-led in its priorities, with built-in arrangements for ensuring full consultation between all interested parties at all stages of the research.

All users contribute to the drawing up of an agreed specification, to evaluation of competitive tenders received, to awarding a contract and to supervising its

execution. The relationship between the JIC and the chosen research contractor is then the foundation for developing and maintaining a professional service for a fixed term, based on a specification drawn up by all prospective users of the system.

Legal constraints, market conditions and cost considerations will determine the form of organization chosen. Likewise, the financing arrangements will have to be determined by each country in the light of its own circumstances. In most countries it is the convention to date for the television networks themselves to bear a majority of the costs of the basic research, but it is often in the interests of the advertising community to contribute. Alternatively, user charges may be levied on the advertiser and agency.

These organizational and funding arrangements are all aimed at striking a proper balance between the responsiveness and the independence of the audience measurement service. Whatever the organizational arrangement, it is a continuing challenge to achieve an audience measurement system which is both fair to all participants (across all sellers, across all buyers and between buyers and sellers) and open to all participants (through appropriate disclosure, discussion and participation).

Whatever type of organization and funding arrangements are adopted to fulfil the Principles outlined above, it is recommended that:

- Formal procedures should be adopted to ensure that the service satisfactorily meets user needs. **User representative groups should have an effective say over the service and the data which are provided.**
- **All aspects of research methods must be open to inspection.** It is vital to user confidence that research methods be transparent. This may be taken care of partly through the specification of quality control and validity procedures as well as through independent audit and accreditation procedures. Only by such transparency can users properly evaluate what is going on and check independently the performance of the data supplier.
- **Different user groups should share the same conditions of access for data used for trading commercial airtime.** Such unrestricted access contributes to the openness of the system, guarantees even trading conditions in terms of audience information between buyers and sellers and facilitates maximum exploitation of the data.
- **Data suppliers should ensure that their survey methods conform to all relevant international codes of conduct.**

DATA COLLECTION

This section addresses the principles and detail of Best Research Practices in data collection, and is followed by the equivalent in the next section on data reporting.

The publication of a formal **Reference Manual** is highly desirable to enable users to establish the precise conventions used in a particular market for each element of its data collection and reporting procedures.

Procedural guidelines are required for a number of different research methodologies:

- **Peplemeter Panels**

This has become the de facto standard audience measurement system throughout the more developed TV markets around the world. The text is therefore directed in the first instance to peplemeter systems. However, the expansion in the number of TV channels and the arrival of digital television in a number of countries are fragmenting the TV audience. These developments are creating demand for larger sample sizes – that may present cost-benefit challenges for relatively expensive peplemeter systems. It is likely that there will continue to be a need for alternative and economical ways of measuring TV audiences – in particular the diary.

- **Diaries**

A number of countries and markets use self-completion diaries, either covering a span of (usually) 7 days or on a continuing panel basis. In either case these systems can be continuous or intermittent.

Diaries are usually deployed where investment in a peplemeter system is not realistic for the local market. But they can also usefully be deployed to meet more narrowly focused information requirements such as:

- **Non-Domestic Viewing**
- **Holiday Homes**
- **International and Niche TV Channels**

Where a diary system requires a substantively different operational procedure from peplemeters, different guidelines are required. These have been highlighted by the convention: **diary**

- **Other Systems**

Other systems include set metering, recall (either face-to-face or over the telephone), the coincidental interview and observation.

The coincidental technique is a specific component of quality control (see VI.20). Use of these alternative techniques is also covered in the sections on Viewing outside the Home and International and Niche TV Channels (see VI.25/26).

1. The Universe

The research universe should be all individuals down to the age of 4 (or a lower specified age) living in private households, without distinction of race, language or nationality. In peplemeter systems this is achieved by installation of peplemeter equipment in a sample of households with at least one television set, to deliver a sample of the universe of the private household population.

Formal definitions of a “private household” unfortunately differ from one country to another. In some it is a person or a group of people living together at one address usually but not always members of one family, whose food and other household expenses are managed as one unit. In others it requires a separate entrance for the accommodation and its own cooking facilities. The universe excludes the institutional population, such as residents of hotels and homes for elderly people, university campuses and the prison population. It should also exclude those who are professionally involved in the industry.

The geographical representation of the universe should be the entire geography of the nation or market to which the audience data are intended to relate. Ideally, therefore, there should be no exclusions. In practice, however, it is rarely cost-effective to cover second homes and holiday homes in the basic continuous audience measurement system. However, the industry may require intermittent coverage of this sector if it is sufficiently important commercially for the market in question.

However, there may be circumstances for deliberately omitting certain sectors of the population, for example:

- **Geography.** There are many territories of the world that are simply impractical to measure because of very low population density or rugged terrain.
- **Danger.** There could be areas too dangerous to research.
- **Low socio-economic groups.** There remain parts of the world where the lower socio-economic strata of the population are not measured, simply because they are not currently part of the viable consumer economy. However this reason for exclusion is rarely legitimate. Television penetration and usage of basic consumer goods are becoming ubiquitous throughout even the lowest income groups. Furthermore, TV viewing varies greatly by socio-economics. Failure to measure the lower socio-economic sector of the population can lead to inaccurate and misleading estimates of TV viewing in a market.

- **Ethnic minorities.** While the research universe should be the total private household population without distinction of race, language or nationality, there can occasionally be exceptional circumstances in which it is simply operationally unrealistic to include a particular ethnic minority who (for example) account for less than 1% of the population and do not speak the language of that country. The data supplier should make any such exclusion transparent to users of the service.

In summary, while the ideal universe is the entire private household population, there can be legitimate reasons to exclude certain precisely defined sectors of the population from the system. Such omissions (e.g. households without a telephone – but see VI.7) must, however, be explicitly stated and justified on grounds of cost or practicality, and any such excluded populations must be clearly and quantitatively defined and excluded from the population projections.

2. The Issue of Quality

Given the importance of audience data to senior management in broadcasting organizations and for the trading of commercial airtime, it is crucial that any audience measurement systems be conducted to the **highest standards possible**. There are two fundamental prerequisites to the fulfilment of this objective:

- First, obtaining the **highest quality sample** of households and of individuals from whom information is collected.
- Secondly, obtaining the **highest quality information** from this sample.

Neither objective is easily fulfilled. Both must be striven for.

3. Towards a High Quality Sample: Probability Sampling

It is highly desirable that any audience measurement system should be based on **probability principles**. A probability sample alone provides projectable population estimates. It requires households to be selected randomly with equal or known probabilities from a sampling frame that, as a reflection of the universe to be measured, is as complete and free of biases as possible (see VI.4).

A quota sample, by contrast, would simply require the sample to conform to predetermined quotas of a range of demographic and other characteristics, such as the age of the main shopper and number of TV sets in the home. Quota samples are not projectable, and are not appropriate to the task of measuring TV audiences.

The major advantage of probability sampling is that it provides an estimate of the sample error associated with the ratings it produces. This means that limits can be put around the “projectable population estimates”. Most sample designs for meter panels (whether one-step or two-step) are complex and special methods are required for error

estimation (bootstrapping or replication). These estimation methods can usefully be built into the sample design itself and incorporated upfront.

There are three equally good basic approaches to the selection of a high quality sample of private households. The method appropriate to a particular country or market will depend on local circumstances.

- Single stage systematic random sample
- Multi-stage systematic random sample (see VI.5)
- Establishment survey as initial stage, that is then used as sampling frame for systematic selection of households for the panel itself (see VI.6)

4. The Sampling Frame

The sampling frame from which a sample (of whatever kind) is drawn is as important as the sample itself. The frame ultimately determines the quality of the sample that can be drawn.

The best sampling frame from which to draw the sample will vary from country to country. Candidates include:

- Census of Population files
- Electoral Registers
- Postal directories or files
- Telephone directories or files
- Church registers in some developing regions of the world

or some combination of the above.

Different measurement techniques may naturally lend themselves to certain types of frame. In most countries of the world where peplemeters are used, the frame is either the national Census of Population in that country, or a combination of the Census and an Establishment Survey (see VI.6).

However, in the case of **diaries**, random digit dialling or telephone or postal directories can satisfactorily be used, subject to the coverage and quality of such government-derived or commercial sampling frames.

The principles governing the quality of a sampling frame are threefold:

- **Coverage** – Is it complete?
- **Accuracy** – Is it accurate – as error-free as possible?
- **Recency** – Is it up-to-date?

It should however, be recognised that there remain countries where real practical difficulties exist in conducting probability surveys. Lists or maps from which samples

can be drawn may not exist. In such circumstances a strict probability sample may not be possible. The research organization should then document its sampling procedures in full, showing how the technique adopted provides the closest, practical approximation to a probability-based sample.

The subject of sampling is complex. Several texts provide useful background and recommended approaches.

Deming. *Sample Design and Business Research* (500 pp.)

Cochran. *Sampling Techniques* (400 pp.)

Kish. *Survey Sampling* (600 pp.)

Dixon and Massey. *Introduction to Statistical Analysis* (500 pp.)

Hansen, Hurwitz and Madow. *Sample Survey Methods and Theory* (2 vol.)

Yule and Kendall. *Introduction to the Theory of Statistics*. (700 pp.)

5. The Multi-Stage Sample Design – and Stratification

En route to the systematic, random selection of particular households whose participation is going to be sought, a number of sampling stages are likely to be required. They can be complex. But fundamentally, what is required is first the systematic selection of a random number of sampling points (often called the “Primary Sampling Units” or PSUs), and secondly the selection of individual households within those PSUs.

The PSUs can be, and are, usefully selected with probability proportionate to size (PPS), that is with probability proportionate to the number of households they contain. An equal number of households is then drawn for each. The sample is in this way inherently self-weighting.

Stratification will usually be built into the selection procedures of PSUs. This may involve, for example, listing a country’s “districts” within its “regions”. These districts may further be ordered sequentially by metropolitan vs other urban vs rural areas and by population size, thus:

- Region 1
 - Metropolitan districts in descending order of population.
 - Other urban districts in descending order of population.
 - Rural districts in descending order of population
- Region 2
 - etc.
 - etc.

In this way it is possible to ensure a good geographical balance in the sample point selection – with an improvement in, what is called, the effective sample size. Stratification procedures lead to less sampling variability than would otherwise occur.

The second stage of sampling is to select a random starting point and then to select every “nth” address down the list, using a fixed sampling interval in the selection process. In this way an unbiased sample of households is selected. Further, the selection has been taken out of the hands of the recruiter. The households eligible for recruitment are determined by the systematic selection procedure.

The number of PSUs should ideally not be less than the number of households planned for the peplemeter panel itself. This procedure ensures that no clustering of households is built into the design of the panel.

6. Establishment Surveys

Some markets find it helpful to conduct Establishment Surveys.

.1 Their Functions and Purposes

Whether or not the market can usefully base its audience measurement system on a high quality and tailor-made “Establishment Survey”, observing strict probability sampling procedures, is a key early decision. Its objectives would be to:

- Produce universe estimates of household demographic and TV equipment ownership characteristics, at the initial stage of setting up a peplemeter system.
- Monitor changes in these characteristics over time, for projection of the measurement system’s universe estimates.
- Provide a sampling frame of households from which the meter panel households can be drawn. Those invited to participate on the meter panel can be selected from those interviewed in the Establishment Survey. Their demographic and TV equipment characteristics are known.

Any Establishment Survey should ideally be conducted annually at the same time of year – or on a continuous rolling basis, thereby ensuring that it is up-to-date. It involves establishing the precise demographic characteristics and TV ownership characteristics of television households. For countries that deploy an Establishment Survey, it is the fundamental and critical means by which changes in the television environment are measured from year to year. The panel of TV homes is adjusted to accord as closely as possible to the latest population estimates of private TV households, as generated by the Establishment Survey.

The deployment of an Establishment Survey is not universal. The industry in any individual market needs to consider carefully whether its requirements of an audience

measurement system can be satisfactorily fulfilled relying solely on government sources (usually the Census) and other high quality published sources.

An Establishment Survey is particularly useful if no other high quality source exists to:

- Obtain universe estimates for specific populations, for example:
 - The geographical coverage of particular stations or cable systems.
 - Multi-set ownership.
 - Video cassette recorder (VCR) ownership.
 - Source of television reception: antenna, cable or direct to home (DTH, “dish”).
- Incorporate a range of panel control procedures for the selection of panel homes, deploying variables related to TV viewing that are not available from any other source (see VI.8).
- Thereby to anticipate and minimise exposure of the system to differential response and potential bias (see VI.7).

There is, however, an alternative to the Establishment Survey procedure. This is to recruit the systematic random sample of private TV households direct from a publicly available, high quality sampling frame. With this procedure there is only one stage in the sampling and recruitment process, as opposed to two phases if an Establishment Survey is deployed – with consequently two opportunities to encounter a refusal. The sampling frame for a one stage area probability sample will usually be based on Census data, the latest published revisions to Census-based population estimates and other high quality sources accepted by the industry as authoritative.

So the decision whether to conduct an Establishment Survey or to employ a single stage selection procedure is a complex one. It must, first, reflect the adequacy or otherwise of the country’s Census data to deliver high quality estimates of the market’s demography. Secondly, it needs to take account of the range of universe estimates required for an audience measurement system some or many of which may not be available from Census sources. An example would be the demography of TV households receiving satellite channels via cable, DTH or SMATV in a particular market (see VII.1).

The decision whether or not to conduct an Establishment Survey to complement Census data should be a joint industry one. The reasons should be fully and quantitatively documented.

For **diary** operations the need for an Establishment Survey does not usually arise.

.2 Questionnaire Content

Information to be obtained from the Establishment Survey, and hence available for determination of the homes to be selected for the meter panel, is likely to need to include:

- Number and location of operational TV sets in the home and in which rooms, distinguishing battery and portable sets.
- Some broad measure of claimed weight of viewing (or equivalent), to distinguish heavy vs medium vs light viewing households (see VI .8).
- Possession of a remote control, VCR, teletext, video games, videodisk or any other device connected to the TV screen.
- All TV channels, both domestic and foreign, to which each set is tuned whether or not they are viewed, and ideally a measure of quality of reception for each.
- Mode of TV reception: antenna, cable, satellite dish etc.
- TV reception: whether in stereo, size of screen etc.
- Telephone ownership
- PC ownership, and whether with or without modem and Internet connection
- Second or holiday home ownership.

If data on specific channels received and/or quality of reception is desired, this should be collected by physical inspection (“scrolling”) of the home’s TV sets. Information on viewing to specific channels is of course subject to reporting error, but may be desired in certain circumstances as a control variable for the panel (see VI.8). In such cases tests should be conducted to ensure that such procedures do not disadvantage any channel or class of channels.

As much detail as possible, **consistent with obtaining the highest possible response rate**, should be collected about ownership of TV related equipment. This should take place whether or not the panel is subsequently controlled by such characteristics, and whether or not viewing is subsequently measured in such terms. The Establishment Survey is a key source of trend data on ownership of a wide range of new emerging TV-related equipment. It needs to achieve, therefore, the **highest possible response rate**: with ideally the in-tab households comprising, if at all possible, 75% of the originally issued and eligible sample.

For countries using a single stage sampling and recruitment procedure (and not deploying an earlier Establishment Survey phase), such TV-related information needs to be collected at the recruitment interview for each panel household. In either event information about each panel household, both its composition and its TV-related equipment, needs to be regularly updated.

7. Panel Recruitment, Maximizing Response, Response Rates

The next task for a peoplemeter system is the recruitment, equipping and training of a probability sample of TV households to participate. A few major problems need to be recognised from the outset.

First, it is likely that the sample size that can realistically be funded will **not be large enough** to meet all user needs. To offset this limitation, clustering of households should be minimised. The ideal is to draw a single stage random sample of addresses, from whatever sampling frame has been chosen for the system – for many countries the Establishment Survey database. A totally unclustered sample of panelists will improve upon the efficiency of the sample design. It follows that the peoplemeter sample design should have as many PSUs as there will be peoplemeter panel households.

Secondly, **high response rates will be difficult to achieve**. The selected household has initially to agree to take part. All members of the household need to be reassured that the metering equipment cannot interfere with or damage their TV, VCR, cable or satellite reception. They may be worried about disruption to the home brought about by installing the equipment. They are then required to meet demanding standards of participation. Each household member is required to register on the push-button handset when they as individuals start and stop watching TV (see VI. 13). Pressing one's button has to become a routine discipline – for each member of the household – from the young child to the elderly grandparent, and whether or not watching TV is the principal activity at the time. The peoplemeter panelist is being asked to fulfil a demanding task.

Thirdly, **high response rates will also be expensive to achieve**. Many more attempts may be required to recruit the less enthusiastic household. Given a fixed budget for an audience measurement system, there is a degree of conflict between maximizing response and maximizing sample size. Priority should usually be given to response. A well balanced panel will be preferable to a larger panel whose composition is subject to significant distortion.

High levels of continuing response are therefore difficult to achieve. Nor is there a single, acceptable target that is applicable and practicable across all countries.

It is general survey research experience that face-to-face interviewing usually achieves higher co-operation rates than the alternatives available such as telephone interviewing. It is the ideal, therefore, for both the Establishment Survey (if one is employed) and the panel household recruitment to be conducted by face-to-face interview rather than over the telephone.

The achievement of the **highest possible response rates** is critical and a necessary condition for the delivery of high quality data. The lower the response rate, the more exposed is the system to the possibility of **bias**. It is commonly found that certain types of family or individual are more difficult to recruit (and also from whom to obtain high standards of compliance to the system) – for example, often:

- Young people
- The elderly
- Light viewers
- Multi-set owners
- High socio-economic status

Audience measurement systems should be closely inspected for evidence of **differential non-response**, leading to systematic under-representation of certain sectors of the population and over-representation of others. In practice, the methods of substituting for households that refuse to participate should be monitored to ensure that they maintain the characteristics of the original sample and do not themselves introduce a bias.

There is, unfortunately, at the moment no universally agreed, formal definition of the crucial concept of **response rate**. Precisely what is, and what is not, included in both the numerator (the number of households providing in-tab data on an average day) and the denominator (the total number of eligible households) of the definition is critical.

Some markets use household recruitment procedures which involve issuing predesignated addresses (the “basics”) and to substitute for refusals (the “alternates”). A common response measure in such countries is:

- The number of “basics” in the active panel providing in-tab data on the average day, as a proportion of the total predesignated sample of “basic” TV households.

This measure of response rate applies only to the data produced from the predesignated households. Where alternate households have contributed to tabulated data, the extent of their inclusion and a complete explanation of how and why they are used should be provided by the research supplier.

Other markets do not employ substitutions for non-participating households, and the measure of response rate applies to all tabulated data. For example:

- The number of households in the in-tab reporting panel sample on the average day, as a proportion of the household addresses originally contacted (net of ineligible).

If an Establishment Survey is conducted, such a response rate excludes non-response to the Establishment Survey itself. This needs to be provided by the research supplier.

In either case, the concept of a “response rate” needs to take account of progressive attrition (see VI.19):

- Refusal at original recruitment:
- Subsequent refusal on installation of the TV monitoring equipment.
- Moving home.
- Voluntary resignation over time.
- Enforced resignation arising, for example, from a panel household consistently failing to meet push-button standards of compliance.
- The proportion of panel homes whose data, for whatever reason, are missing from the average day’s in-tab database.

The scope for **bias** in TV audience data is clearly considerable if, as the European experience suggests, the number of households in the in-tab reporting panel sample on the average day **is about one in three of the original households contacted for recruitment**. The two-thirds that have not fully co-operated, in one sense or another, could be materially different in their TV viewing from the one-third conscientiously serving on the panel. For countries therefore using an Establishment Survey (with response rates typically about 75%) , the real (longitudinal) response rate is likely to be as low as 25%: 33% of 75%.

Response rates can be significantly improved by careful attention to recruitment techniques, including persistence, allowing time to convert refusers, employing representatives trained in door-to-door selling techniques, etc.

The industry is strongly encouraged to work towards a standardized definition of “response rate” and towards achieving the highest response rates possible.

It will be seen that, however defined, there will in all likelihood remain a considerable and worrying scope for bias in the composition of the panel. It may, for example, be more difficult to recruit and obtain the effective co-operation of wealthy sectors of society of particular importance to the advertiser, and of large families with children and several TV sets. In such circumstances the panel could so easily, unintentionally but systematically, be biased towards the lower socio-economic strata and towards smaller households. Young one-person households are particularly difficult to recruit. There should not, however, be a problem in representation of non-telephone owning households. They can if necessary be equipped with radio telephones for daily downloading of viewing data.

A distinction needs to be made between differential response rates amongst various geo-demographic segments and differential response rates in terms of viewing behaviour. One does not imply the other. It is fairly easy to re-weight data to geo-demographic population distributions, and to observe whether there is any resultant change in viewing estimates. Beyond that, there can be differences in viewing

behaviour between responders and non-responders which will distort audience estimates, even though the sample may be perfectly balanced in terms of its geo-demography. This is the fundamental reason why the need to maximise the response rate (and to conduct non-response research) is paramount.

It is also important to measure all TV viewing in multi-set households. Metering technology requires continuing development to achieve this. Where in a prospective panel household there exists one (or more than one) set that cannot technically be metered, this household should not be recruited if this/these appear to account for 5% or more of the total household's TV viewing. Some broad information is therefore required at the recruitment interview to establish this.

The issue, therefore, of precisely how the addresses (leading to households) are selected for recruitment matters a great deal. Statistical theory leads to one and only one correct solution. The household units should be selected in accordance with systematic probability procedures. However, statistical theory has no ready-made answer for the problem of **bias**. Solely to rely on systematic random selection does not protect a system with a response rate (in absolute terms) on the low side from bias. One helpful solution to reduce exposure to bias is the deployment of **panel control procedures** (see VI.8).

For recruitment of **diary** samples or panels, again strict probability sampling is the theoretical ideal. However, again there are likely to be response rate problems. A significant task is being required of the respondent: namely to keep a diary of his or her TV viewing conscientiously over a span of (for example) 7 days – or, in the case of a diary panel, on a continuing basis. The task of keeping the diary should be kept as simple as possible. The collection of demographic and other classification data should be limited to the strictly pertinent – in the interest again of maximising response to the study.

Further, the research budget the market can afford may not permit personal placement and collection of the diary. If telephones are owned by the vast majority of households, it may be decided to conduct the sign-up interview over the phone, to be followed by postal delivery of the diary and reliance on the respondent to remember to post it back at the end of the week. The effective response rate is likely to give scope for considerable bias in the achieved sample.

The principles of:

- Coverage
- Accuracy
- Recency

apply equally to **diary** sampling frames as for peplemeters. It is equally critical to use

the most cost-effective route towards obtaining the **highest response rate** practicable. This is a highly complex issue, in principle involving assessing specific combinations of:

- Face-to-face vs telephone recruitment
- Personal vs postal placement
- Personal vs postal collection

8. Panel Controls

One technique to reduce exposure to bias is to **include a range of panel control procedures**. They will offset the scope for significant bias being unwittingly built into the originally signed up panel, and progressively worsened over the life of the panel. The technique is to deploy selected items of household information as controls for the original homes selected for recruitment and for panel replacement thereafter.

The ideal selection of panel controls will vary from country to country. The ones chosen should be related to viewing behaviour. If the final recruited panel closely reflects the universe in terms of these characteristics, then bias in terms of these control variables will have been significantly reduced, and one condition for the design of an unbiased TV audience measurement system will have been fulfilled. Panel control variables that have been found to be the most valuable often include:

- Number of persons in household: large vs small.
- Presence of children in the household.
- Number of TV sets.
- Channel reception, to provide for correct representation of channels with varying geographical penetration.
- Age of principal wage earner, or equivalent life-stage variable.
- A socio-economic variable.
- Ethnic community representation (where relevant for the country).
- Language(s) spoken (where relevant for the country).
- Mode of television reception via cable and satellite, or over the air.

The source for these population estimates will in part be Census-based and in part drawn from the Establishment Survey (if one is employed). It should be remembered that Establishment Surveys are themselves subject to non-response and response biases. As has been pointed out above (see VI.7), **panel controls will not deal with whatever systematic differences in viewing behaviour may exist between responders and non-responders** – for which the only solution is the maximisation of response rates. So when both sources are available, panel controls should normally be based on Census data rather than on an Establishment Survey. Individual addresses need then to be selected in a totally systematic way from whatever sampling frame is being used to meet specific household-type criteria.

Such panel control procedures are quite **distinct from quota sampling**. Quota sampling would leave the choice of the household in the hands of the recruiter. By contrast these procedures ensure that addresses issued (leading to households) are predetermined and drawn in a systematic way from whatever sampling frame is being used in the country. The selection of households for recruitment has been taken out of the hands of the recruiter.

There is one particular procedure that can, if correctly handled, be especially useful in minimising a system's exposure to **bias against light viewers**.

It is the experience of most countries that light viewing households are more difficult to recruit to a peplemeter panel. They are less interested in television. It follows that, if no counter-measures are taken, the structure of the recruited panel is likely to be biased against light viewers. The outcome will be an over-estimate of TV viewing.

This quasi-stratification procedure, that can only be adopted in countries using an Establishment Survey, involves asking the respondent (usually the Housewife or Main Shopper) for an estimate of the total number of hours each TV set in the household is switched on across the average weekday, Saturday and Sunday. Households are then listed in descending order of TV set hours. The sample is divided into three equally sized groups to represent heavy, medium and light (or non-) viewers respectively. The panel is then recruited in such a way as to ensure that one-third of the panel is randomly drawn from each of these three strata in a systematic way. The panel is being balanced by **claimed weight of viewing**. It is a discipline similar to stratification (see VI.5).

Similarly, in a market dominated for example by two channels it may be helpful to establish in broad terms the relative time the family spends viewing each. Tritiles can then be established in terms of the ratio of viewing Channel A to Channel B, and used at the recruitment stage to ensure that the recruited panel is not unwittingly biased in favour of one of the two channels.

A criticism sometimes made of such procedures is that they would appear to be controlling the panel's composition by the very variable that it is the purpose of the panel to measure: namely TV viewing. However, this is a misunderstanding. These procedures do not require the actual viewing behaviour as measured in due course by the peplemeters to be what was claimed in the original Establishment Survey interview. It would be wrong to remove a particular household from the panel where a substantial discrepancy was found between claimed and actual. The outcome would indeed be panel selection by the very variable one is setting out to measure. All that is required is some reasonable correlation between claimed and actual. Otherwise this stratification procedure would not be achieving any purpose.

By analogy, for a survey on household income it would be efficient and appropriate to stratify the sample by known but less detailed information on income drawn from the national Census.

9. Sample Size

The sample size of the panel of TV homes should be determined by the precision required. This is likely to be set by the range of demographic and geographic population sub-groups whose viewing is required to be reported, for commercial trading or editorial purposes, and by the number/size of channels to be reported. It should also take account of the efficiency or otherwise of the sample design as determined by the sampling method chosen (see VI.5 above).

The overall panel sample size for a country is often therefore the aggregate of these requirements. They can vary accordingly from around 5,000 TV panel homes in a large country to around 500 in a small market. It is unlikely that any country (or major city) could require a sample size less than 300, net of non-response to the measurement system.

In all but the most limited media markets, the bigger the sample the better. A minimum sample is generally considered to be in the 300 range, net of non-response. "Better" would be in the range of 1,000, enabling finer demographic groups to be analysed, while "best," based on current practice, would be 5,000 or more. Markets are strongly encouraged to move to larger sample sizes as the number of channels to be measured increases, and as funding permits. It should always be kept in mind, however, that size is no substitute for a carefully and systematically drawn sample, based on as high a response rate as is realistically achievable, and incorporating tried and tested techniques to minimise the system's exposure to bias.

A disproportionate sample design may deliberately be chosen. It may be decided, for example, to over-represent households in a particular geographical region, social class (SES) or TV reception mode by a factor of (for example) $\times 2$, to obtain improved estimates of TV viewing in such households. Corrective weights are then applied, in this case of $\times 0.5$, to return the sample to one that is projectable to the country as a whole. (But see the cautionary note about weighting in the final paragraph of VI.22.)

Appropriate sample sizes for **diary** systems will again depend on the precision required and the range of demographic sub-groups to be reported. However, it is unlikely that a market would be reported on a sample size of less than 500 households, involving either the random selection of one person per household, or "flooding" where all household members are asked to participate (see VI.17). Corrective weighting procedures will be required for single person diaries, if the households have been selected with equal probability, since otherwise (for example) individuals in three-

person households will only have one third the probability of being chosen relative to individuals living alone.

10. Panel Incentives

Achieving the highest response rate possible is as much an art as a science. Financial incentives will nearly always be involved and may for example include a contribution to the household's telephone rental or prize draws.

A regular newsletter may also be found effective. The purpose of a newsletter will in part be to emphasise the importance of the panelist's role, in contributing to accurate measurement of their television viewing. It is likely to be used to communicate certain aspects of what is required of a panel member, whether for clarification or to improve standards of compliance.

All incentives should be made explicit. They need to be cost-effective and to depend on the conditions of each country. Experimentation with a range of incentive strategies is recommended.

Care should be taken to ensure that the incentive scheme does not itself influence viewing behaviour. For example, it would be inappropriate to offer a free TV repair service. The outcome would be an artificial reduction in the number of TV sets out of order which otherwise would not have been watched. In general, large incentives should be avoided. Large sums of money (or the equivalent in goods) risk changing respondent behaviour – for example, in the entertainment goods and services they subsequently acquire, or in the kinds of holidays they can afford to take.

Undue reliance should not be made upon incentives. The commitment by panelists to take part in an audience measurement system is an act of generosity on their part. The incentives must not belittle that altruistic spirit, but be offered as reciprocal acts of generosity. If incentives are portrayed as compensation for the effort put into achieving high standards of compliance with the system, it is so easy to insult panelists by the small size of the incentive relative to the demands on them of panel participation.

11. Panel Maintenance and Renewal

Panel turnover arises from six principal causes:

- Households moving.
- Voluntary resignation from the panel.
- Poor standard of panel performance leading to enforced resignation.
- Enforced resignation to achieve panel balancing to control criteria.
- Enforced resignation if the identity of a panel home becomes public.
- Any maximum length of panel membership that the system may set (see below).

The aggregate panel turnover arising from all these elements in total should be estimated and budgeted for realistically. This will vary from one country to another and will probably be in the range of 15-25% per annum, excluding and before the requirements of any enforced turnover procedure.

The need to consider setting a maximum length of panel membership arises from the possibilities that (a) push-button behaviour deteriorates over time, or (b) sample bias is introduced by the operation of extraneous variables, which influence individual length of panel membership. The principal argument for incorporating an enforced turnover system is to reduce both fatigue and ageing of the panel. Although enforced turnover after as little as six months has sometimes been employed, maximum in-panel periods of one year, two years and five years are the most common. Each market should study its own fatigue and ageing patterns to determine the most appropriate turnover period.

There are pros and cons with enforced turnover. The benefits of enforced turnover have to be weighed against:

- The potential reduction in sample sizes available for longitudinal analyses.
- Potentially less stability in trend data.
- The extra costs involved in operating an enforced rotation system.

Similar pros and cons are to be found with other forms of enforced turnover, such as periodically discarding selected panelists in order to maintain a constant average length of service on the panel, thereby to improve panel stability. In this instance, there would also have to be a transition period between the initial recruitment of panelists, when all households are new to the panel, and the eventual stable situation, where the composition of the panel in terms of length of membership is stable. Decisions that would have to be taken include the length of the transition period; the definition of the length of service targets; the management of the transition stage; and the procedures for maintaining thereafter a stable panel profile by length of service.

No matter what approach is being considered, **monitoring of panel performance remains essential**. No fixed procedure of enforced turnover is a proper substitute for the regular monitoring of panel performance. Panel performance should be regularly monitored to look for evidence of panel fatigue, by examining trend data in TV viewing patterns by demographic control groups and by the period already served on the panel. For example, is there a systematic tendency for the amount of teenager viewing registered by the system to be higher amongst those teenagers who are in their first 6 months of participation relative to those who have served for at least 18 months? If this is found to exist, then some systematic procedure of enforced turnover may be desirable.

12. TV Audiences to be Measured

The system should measure:

- The **total (live) audience** (see VI.13 for definitions of “audience”) to all channels and programs across all TV sets in the home, over the 24-hour day, with use of the VCR as a tuner counted as live viewing. (Digital television is posing challenges for data suppliers in this respect.)
- The total audience contributed by **guests**.
- The audience to **all sets** in the home that can be technically and cost-effectively metered. Otherwise there is a risk of bias in the measurement system, for example against proper representation of the viewing of particular channels (see VI.7). Portable and battery sets may not be practicable. There may also be technical problems in metering certain specific sets in the home.
- All other uses of the TV sets, for example teletext, viewing of rented or purchased videos, PC-linked uses, video games, radio listening etc.

The market may or may not require measurement of **timeshift viewing**. This requires VCR encoding of channel, day and time of live transmission, to be decoded at playback.

There is an important distinction between what should be **measured** and what is needed for **reporting**. All forms of TV usage should in principle be measured, so that they are capable of being reported whether or not a particular form of viewing (for example teletext usage) is required for routine reporting.

For TV audience measurement using the **personal** individual **diary**, the diarist’s own out-of-home viewing should be registered. There is therefore no requirement for guest viewing as a surrogate for this. For **household** diaries, however, it is often the convention for a diary to be issued for each set in the home. The diaries are therefore set-specific, and presence of guests in the audience therefore needs to be registered (see VI.17).

Diary registration of time-shift viewing poses some practical problems in establishing the channel and program being viewed after the event. It may be decided to limit diary systems to the live audience.

13. The Definition of “Audience”

It is one thing to say that the total live audience is to be reported. It is another to say what this is. There are at present a number of conceptually different definitions in use – whether for peplemeter systems or for diaries:

- **Presence** in the room – with the TV set switched on.

- Respondents registering when they regard themselves as watching television: a **subjective** definition.

Whichever definition is used should be carefully explained to panelists at the time when the household's active membership of the panel begins. If they are instructed to register their presence (or record their viewing in a **diary**) simply when they are in the room, then the first definition becomes the de facto definition of audience. Alternatively they may be instructed to press their button (or enter viewing in their **diary**) when they regard themselves as watching, in which case the cognitive definition is being adopted.

Some countries have adopted a definition that lies between the above two extremes:

- In the room and able to view

There is some experimental evidence, however, to suggest that what is being measured under each definition of viewing is not very different. This may be because panelists impose their own psychological interpretation of when they count as viewers, regardless of the actual wording of their instructions. For example, where "presence" alone is the criterion, they may tend not to regard themselves as watching television while engrossed in another activity (for example reading a book or newspaper) and hence not to register themselves as viewers in spite of instructions to the contrary. However, strenuous efforts should be made to be as specific and clear as possible to panelists on what constitutes counting themselves as in the audience. Inadequate instructions are likely to exacerbate the tendency for panelists to impose their own definitions.

Each of the definitions has its advantages and its disadvantages. In the case of the "presence" definition the interpretation of results of coincidental checks carried out amongst samples of panelists is straightforward (see VI.20). The coincidental information being sought is the room in which each member of the family was located at the precise moment of the check call. It is established whether they were "present" in a room with a TV set on. This is checked against whether or not they had registered their "presence" by pressing their individual button on the system's handset. Secondly, in practice individuals do not always necessarily press their own button. Other members of the family may be doing the button-pressing on behalf of young children or guests. The basic "presence" criterion is the most practical one for such circumstances. Thirdly, should the era of passive sensing become a practical cost effective reality, this is likely to involve a concept of viewing close to that of "presence".

However, the notion of a conscious state of TV viewing that is nominally delivered by the subjective definition is intrinsically attractive to both program makers and advertisers. Further, there are circumstances, such as the viewing of communal sets in developing countries, where the subjective definition has obvious practical advantages.

In the circumstances, there is no universally recommended definition. Each country or market should decide on the definition that best meets national circumstances. The rationale behind that decision should be explained.

14. The Peplemeter Technology

In the case of current push-button technology, it is recommended that the equipment should be user-friendly and cater for:

- As many household members as is necessary and practical for the market.
- As many guests as is appropriate for the country and is practical. (The development of pay-per-view is likely to lead to the guest component of the TV audience becoming increasingly important.)
- As many channels as can be received in multi-channel homes in that market.

Its design should be such that it should not influence or modify panelists' viewing behaviour. For example, use of the conventional remote control needs to be unaffected by participation in a peplemeter system, and should not be replaced by the peplemeter handset.

It should be suitable for all ages: from young children to the elderly. It may also be helpful to incorporate a systematic reminder signal to ensure that both household members and guests register their personal presence in the audience. It should, however, be recognised that this poses an asymmetrical reinforcement. A reminder signal can prompt panelists to log in. It cannot prompt them to log out when they leave the audience. It can be useful for the handset to have a "no-one in the audience" button to indicate occasions when the TV set is switched on but no-one is there. This condition legitimately arises, for example, when the total family leaves the TV on, to deter burglars or to keep a child under the age of 4 entertained. In such circumstances no one (aged 4 and over) should be registered as being in the audience. A "no-one in the audience" button contributes to both quality control and data editing protocols (see VI.16, 18, 21).

The system should be capable of identifying individual seconds of TV use, for both the TV set and push-button status, as the basic unit of measurement – whether or not the software is created to identify each individual second.

15. Guest Viewing

Guests in the home who watch television should be registered as TV viewers. They are an element of the private household's TV audience, and may be differentially important to certain (for example children's) stations. In aggregate their viewing is equivalent to the viewing of panelists at their friends' homes. Guests can usefully be

required to enter their sex and age. One convention is to apply to the guest the socio-demographic classification of the home being visited.

There are, however, practical problems in gaining co-operation from visitors. There is a tendency for guest audiences to be understated by peplemeter measurement systems. The task for guests needs to be as simple as possible. Only the minimum information needed for commercial or editorial purposes should be sought. Other stratagems to improve the registration of guest viewing include minimizing the number of keystrokes required to register their viewing, and allowing regular guests (for example, grandparents who visit the family frequently) to have their personal number on the handset – even though technically they are not members of the household.

Whether guest viewing is routinely included in TV audiences as reported is quite another matter: see VII.3.

Again, principles of transparency in research methods should be followed. How guest viewing is handled should be spelled out in the published reference manual for the system.

Guest viewing does not arise for the individual **diary** but is required for the household diary (see VI.17).

16. Nil Viewing – and Panelists on Holiday

Correct interpretation of various kinds of “nil viewing” is important:

Situation	Interpretation	
• Someone is at home, one or more TV sets are switched on, but no panelist is in the audience	• Information acceptable	✓
• No-one is at home – perhaps on holiday. One or more TV sets may however have been left switched on, for example to deter burglars	• Information acceptable	✓
• This is a power failure, arising perhaps from a thunderstorm.	• Information acceptable	✓
• TV set broken and awaiting repair	• Information acceptable	✓
• Someone is watching television but has not registered their viewing	• Such individuals need to be identified and corrective action taken to improve their compliance (see VI.18)	X

Absence on holiday is in particular a key issue. Such homes should be included in the in-tab sample. To exclude such homes would be to assume, improperly, that panelist viewing on holiday is no different from when they are at home.

In countries with a high incidence of second homes, separate and independent techniques may be required to obtain estimates of viewing in second homes which would otherwise be missing from the TV panel's estimates of total TV viewing in private households (see VI.25).

17. The Diary

There are a number of **diary** designs currently in use around the world. The principal differences between them arise from two design issues:

- Whether the intention is to recruit **individual** diarists – usually one per household, or to attempt to sign up **all household members** down to a predetermined age (for example 4)
- How the **channel** is entered into the diary – and hence how the program being watched is identified

In other respects, for example time unit conventions, there is more standardization.

.1 Time Units

A time unit of 15 minutes is the realistic and satisfactory convention for diaries. To qualify as a viewer for a particular 15 minutes (beginning xx00, xx15, xx30, xx45 hours) the technically correct minimum should be **8 minutes**: more than half the period. In theory a single minute is then zero, and is balanced by 14 minutes being given a value of 15 minutes, etc.

However, there is one quite important difficulty with this eight minute definition. Again in theory at least, a person who only tunes to a channel for (for example) brief news bulletins or the weather forecast (of less than eight minutes' duration) would never be registered as a viewer of that channel. The Reach of specialist news or weather stations could be systematically understated relative to other stations. On the other hand, to allow **any** viewing to qualify would lead without doubt to a significant degree of overstatement of viewing hours. A one minute news bulletin, for example, would be given a value of 15 minutes. A practical compromise is to set **5 minutes** as the minimum qualification.

Another useful convention is to permit diarists to register themselves as viewing more than one channel during any particular 15 minute period within which he/she switches channel. However, steps then have to be taken at the data editing stage to divide up the 15 minutes between the channels in question (see VI.21). TV viewing would be overstated if each channel were credited with 15 minutes of viewing under such circumstances.

All 24 hours of the day should be covered. However, for a diary to keep rigidly to the 15 minute time unit throughout all 24 hours of the day can look daunting to the diarist –

and also be unnecessary in those countries whose principal channels are not on air during the night. It is acceptable to use broader time periods, either of 30 or of 60 minutes, during the night with at least 10/20 minutes being the minimum to qualify as viewing for that period. The precise conventions for broader time registration should be determined by industry consensus for the country or market in question.

.2 The Individual vs Household Diary

The principal options for this key design issue is whether the plan is that the diary should relate to the total viewing of one individual (both in their own home and elsewhere), or whether it should cover the viewing of all household members down to a predetermined minimum age.

This latter technique is sometimes known as “flooding”. One of its attractions is the large sample size of diarists that can be delivered by a given research budget. However, to the extent that people within a household have similar viewing habits this intra-household correlation leads to a somewhat lower effective sample size. (This is analogous to the convention used with traditional set meters. Where diary data are combined with meter data to produce meter-adjusted audience estimates relating to individuals, it is recommended that complete households be measured in both samples.)

For household diaries a satisfactory convention is to provide a diary for each set in the household. A three-set household receives three diaries, each of which is therefore set-specific to a particular room. The viewing of each household member is entered for each set. It follows that the TV viewing being measured is limited to in-home viewing. It is therefore necessary to include guest viewing (see VI.15) as the surrogate for the out-of-home viewing of the diarists themselves.

.3 Channel (and Program) Registration

In deciding how to identify channels in the diary it is of crucial importance to ensure that:

- All channels viewed are recorded.
- No bias is introduced by the measurement system itself in favour of some channels and against others.

One approach is to divide the country or market into as many geographical sectors as necessary to reduce the number of channels lying within each sector down to a manageable number. Channels are then pre-listed in the diary. This is sometimes called the “rostered” diary. There will usually be a need to produce a large number of diary versions, such that the geography of **all significant channels** is fully covered by these variants, **whether or not they are subscribing to the service**. Some channels

would feature in more than one version – these including all national channels of course.

Clearly there needs to be some practical limit on what comprises a “significant channel”. A rigorous set of rules needs to be drawn up in each country on the criteria for channel inclusion, following full consultation with all sectors of the industry: broadcasters, advertisers and agencies. In some countries it may be necessary to list a number of overspill foreign channels. It is recommended that any channel with a Weekly Reach of 5% and above is “significant” and should be identified in the system. Some advance experimentation is then required to establish a complete list of stations that qualify.

There will additionally be some viewing of channels that have not been pre-listed. The diarist can be asked to enter the names of these channels. However, there will inevitably be some degree of understatement of (i.e. bias against) unlisted channels relative to the rostered channels.

Each diary page also needs a code for diarists to register when they did not view television at all during the period. At the analysis stage it is then possible to distinguish nil-viewing from forgetting to make any entry.

There will be, however, many countries where the sheer number of channels available make any form of channel pre-listing procedure particularly difficult. Instead the respondent is relied upon to enter the channel name (and sometimes program name as well) into the diary. This is the “unrostered” diary. New problems now arise:

- This unaided technique will often register less viewing than a technique involving channel pre-listing. Viewing is relatively under-stated.
- Secondly, this relative understatement can disproportionately affect channels rarely viewed. Channels with large numbers of viewers are also usually the channels in the regular repertoire of their viewers. “Minor” channels can have proportionately fewer regular viewers. Paradoxically, the outcome of what appears to be scrupulous even-handedness can be bias.
- Thirdly, there is the problem of precise channel/program identification. The respondent may not have clearly identified it. It may be necessary to get in touch with diarists again over the coming weeks at the diary editing stage to try to establish which particular channel was being viewed on a particular day.

The totally unprompted approach to channel identification therefore also has its difficulties.

In summary, recall of channels viewed is a serious problem, especially in markets with a large number of channels due to widespread cable/satellite, etc. Some type of prompt (such as a diary roster) should be provided wherever possible. Where this is not possible, research should be conducted on diary design and fieldwork techniques to minimize recall bias to the greatest extent possible.

18. Quality Control

Quality control procedures are critical to the production of valid and reliable data. The range of procedures observed should comprise a technical manual in their own right and be transparent to all users. The following are examples of significant quality control. The research institute is encouraged to adopt these and, where possible, adopt others that are applicable to their own specific methodology and market conditions.

- **Fieldwork Verification.** This applies to the entire data collection procedure for **diaries** and to the sample design, selection, installation and maintenance procedures for peplemeter systems, including Establishment Survey or other universe derivation procedures deployed by the research institute. At least 20% of each fieldworker's assignment should be independently monitored and validated.
- **Information System Controls.** All computer programs and software, and changes thereto, used in the processing and delivery of reported results should be authorized, adequately tested and approved prior to being placed into production at the research institute. Computer programs and software should be adequately documented, including changes made with an audit trail. Computer programs, software and data of the research institute should have adequate physical and electronic security measures to prevent access by research institute personnel, or outsiders, without a business need to know.
- **Changes in Household Composition and TV Equipment Ownership.** Panel members should be instructed how to report changes in household composition or in additions or removal of equipment used to receive television. Monitoring of meter data may trigger enquiries about possible changes that have not been reported. In any event, periodic enquiries to reconfirm this information for each panel household are essential to the accuracy of television audience data. Where unmetered equipment is discovered, it is appropriate to withhold that household from the in-tab sample until such time as their equipment is fully metered again.
- **Push-Button Performance.** Monitoring the diligence or otherwise of panelists with their push-buttons is critical. The push-button approach to

measuring TV audiences relies on conscientious registering of viewing by each member of the family (down to the age of 4) and by their guests. To identify at an early stage when individual panel members become lax in their participation is critical to the quality of the data and the validity of the system. Examples of detection procedures which could be used by a research institute include: (1) enquiries following data monitoring, (2) within-panel coincidental exercises, and (3) either internal or independent auditing. Panelists should be encouraged as much as is possible and is reasonable to be committed to accurate and complete recording of their TV usage.

- **Data Editing and Reasonableness Tests.** Formal procedures are required for investigating apparent nil viewers, viewers with extreme viewing patterns and households or individuals that exhibit extended periods of uncovered set viewing (when the set is on but no one has registered their viewing). These procedures should include two steps. First, comprehensive software checks (edits or reasonableness tests) should be used to identify for each day's data all instances of extreme (high, low and nil) and unusual viewing. Secondly, households identified in these software checks should be followed up. It should be noted that unusual viewing activity is sometimes accurate based on household circumstances, and may not require further action. It is critically important that the follow-up performed by the research institute do not lead to conditioning of panelists subsequently to change their viewing behaviour.
- **Monitoring of Panel Characteristics.** The panel needs regular and systematic checking to look for deviations of the key demographic and geographic characteristics of the panel from the universe estimates. If panel controls are used, the panel needs to be continuously monitored to ensure it remains within the tolerance limits of the control characteristics. If other mechanisms are used to reflect migration or other population shifts in the panel, these should be applied at regular intervals and accurately.
- **Data Changes.** Ascription, editing, or data adjustment procedures used by the research institute in the production of reported results should be checked on a periodic basis, and the methodology used should be disclosed to all users.
- **Set Meter Clocks.** Checking the accuracy of the set meter clocks is necessary. Additionally, time adjustments or shifts made by the research institute to reflect time zone changes or multicasting should be periodically validated.
- **Telephones.** The functioning of panel telephones used in the data communication process should be monitored daily. Otherwise, data loss could result.

- **Data Retention.** Significant source documents from the research institute's methodology such as panel records, viewing data, data removed from production due to edits or quality control failure, etc. should be retained for at least 12 months after the applicable ratings report is released. For example, recruitment records of all households used in the survey should be retained for at least 12 months after they leave the panel.
- **Confidentiality.** The names and addresses of respondents must be kept confidential and not become known to users. The confidentiality of the identities of the panel households is paramount to the operation of an independent, unbiased system. Homes that are identified by persons not covered by a confidentiality agreement should be dropped from the panel. Within the research organization, the information should be on a "need to know" basis. Other organizations, such as cable or DTH suppliers, who have to supply or install special equipment, should also be bound by a confidentiality agreement.

Quality control procedures should be systematized and be transparent to the users of the system. Insofar as possible, care should be taken to ensure that these procedures do not create behavioural changes in panel households or a deterioration in panel performance.

Significant errors discovered in the production process of the research institute or in its reported data, should be disclosed to all users.

19. **Monitoring System Performance**

A monthly trend report should be provided by the research institute giving a wide range of information about the performance being achieved by the system. This should include, for **diaries** as appropriate:

The Panel/Sample

- Response rates being achieved for **diaries**, the Establishment Survey and panel recruitment – distinguishing between predesignated and substitute households as appropriate:
 - No reply to any call
 - Address ineligible, distinguishing between different categories of ineligibility (for example a campus lodging, a hotel, premises unoccupied)
 - Individual selected not available

- Refusal
 - Failure to place
 - Failure to receive back
 - Rejected at editing
 - Effective in-tab (see VI.21)
- } – for **diaries**
- The Panel:
 - Average daily panel/sample size:
 - recruited
 - installed
 - reporting
 - in-tab
 - Actual versus target sample sizes for each panel control cell – including mode of reception (cable/satellite etc.).
 - Homes on holiday – while accepting some operational problems in defining “away on holiday”
 - Home reclassifications, for changes in (for example) family composition, equipment ownership:
 - made
 - scheduled or required
 - Drop-out rates from the panel, within:
 - panel control groups
 - length of service on the panel
 - category of panel turnover (see VI.11 above).
 - Metering:
 - Engineer visits
 - Average daily polling response.
 - Proportion of set usage with station and/or viewer unidentified.
 - Proportion of set usage with no individual registered as viewing.
 - Average weekly set non-usage.
 - Viewing:
 - Average hours of viewing.
 - The incidence of extremes of hours of viewing – both low and high – to look for possible meter malfunctions and other errors.
 - Guest viewing as proportion of total.
 - Viewing by mode of reception (cable, satellite etc.)

As such published performance data are developed, the international research community should set itself the task of agreeing upon cost-effective functional performance goals for operating peplemeter systems. **However, any agreement**

between the parties on the minimum standards to be achieved does not remove the onus on suppliers to aspire to the highest possible standards.

20. Coincidental Studies

At least once a year Coincidental Studies should be carried out to monitor the validity of the measurement system and panel performance. The object is to establish whether at particular selected times on particularly selected days the system is delivering a valid measurement of the TV audience.

The Coincidental Study can either be conducted amongst a sample of the panelists themselves or by an entirely independent measurement exercise. The criteria of validity being examined are respectively very different.

The **within-panel** approach of establishing contact, usually by telephone, with peoplemeter panelists at precisely predetermined times has the narrow but very specific objective of establishing quality of push-button performance. In what proportion of instances when panelists are viewing television do they fail to register this? Equally how many instances are found of panelists incorrectly registering themselves as viewing? One is setting out to measure the gross and net effect of errors of push-button registration. Panelist fatigue can be measured by analysing such error rates by length of time on panel. Care needs to be taken that these contacts with panelists are conducted in a manner that is not likely to lead to conditioning of their TV viewing. If described and conducted as part of the ongoing quality control process, periodic within-panel coincidental interviews can serve to reinforce the broad goal of measurement accuracy.

The **external** exercise of obtaining an entirely separate measure of the TV audience at predetermined times on selected days is establishing whether the overall measurement system is delivering audience levels that are correct within the limits of sampling error. Where differences in audience estimates however are found, interpretation may not be straightforward. The possibility of imperfections within the Coincidental Study itself clearly cannot be ruled out.

The internal and external approaches to validation are not substitutes for each other. They respectively provide information about the validity of a peoplemeter system that is different and complementary. The internal investigation examines the important minutiae of panelists' push-button performance. The external validation study sets out to measure overall system validity – providing a separate and independent measure of the outside reality.

While the external coincidental study is often considered the “gold standard” against which to measure the performance of other methodologies, it is subject to its own biases which can be serious and should be carefully estimated and monitored. The

technique is dependent on factors including telephone penetration, completion rates, daypart limitations and sample size (which, in a multi-channel universe, may have to be large). Major coincidental studies should be conducted outside the active panel wherever possible, in order to minimize the effect on reported audiences. Any coincidentals that are conducted within the live panel should be disclosed in the ratings reports which they might in principle affect.

Complementary to the Coincidental Study approach is the use of other independent high quality sources of data.

21. Editing Rules

Formal and systematic editing procedures to correct for imperfect data are necessary – whether in diary-based systems or from peplemeters. These should form a **manual in their own right and be transparent to all users.**

Explicit rules are required for the treatment of **missing data**. Should that day's data for that particular household be included or excluded from the in-tab sample for that day? To include the household assumes **nil** viewing in effect for the item of information missing. To exclude it assumes **average** viewing for the household in question. These are key quality control issues. Explicit rules need to be formalized and agreed with the industry. Development of a methodology for edit rules is not a trivial task. Within-panel coincidentals and examination of normative periods of "good" data from the households in question can contribute greatly to this process.

In the case of peplemeter systems a problem arises when there is a set meter record of the TV being switched on but no one has registered himself as viewing. This can happen at the beginning of a viewing session, at the end of a viewing session, at some point within it, or throughout its entirety. Sharing of validation data amongst research practitioners should be encouraged to understand better the reality behind the "uncovered viewing" gaps, and to reach a consensus on how to treat them and precisely what imputing of viewing into these "gaps" is desirable.

Another less contentious form of data editing is to exclude live viewing of a station outside its transmission hours.

Variations in practice can materially affect estimates of viewing. Some standardization of procedures is therefore essential in order to obtain comparable measures of TV audiences between countries. In any event the precise editing rules being used in any system should be fully declared and published in its reference manual.

Explicit rules need to be drawn up for checking of respondents with extreme values of viewing time.

In the case of **diaries**, the principal edit is to identify quarter hour periods during which more than one channel is watched and to divide the 15 minutes attributed viewing evenly between them – this being the fairest convention to use. More precision in the registering of starts and stops in TV viewing is probably unrealistic. The procedure ensures that total daily viewing time is correctly calculated.

In addition the problem arises of how to treat **diaries** that appear to be fully and conscientiously completed, but the odd page is blank. Is the correct interpretation that no viewing took place and the respondent simply failed to register this explicitly by coding the nil-viewing box? Or that some TV viewing did take place and has not been entered? One convention is to accept the diary is valid providing that at least the first and final days of the diary-keeping week are completed and at least two (or three) other days. Blank days are then interpreted as days on which no TV viewing took place. As with meters, research that examines in detail how diary keepers maintain their diaries can provide a factual basis on which to base edit rules.

Care should be taken to ensure that diarists who do not view television during the survey period are identified as such and are included in the database.

22. Weighting

The purpose of applying weighting to survey data is to restore it to agreed population projections, but it will not correct for all the biases arising out of differences in TV viewing behaviour between those households that agree to participate in a peplemeter system and those who do not. The need for weighting can also arise from the inherent design of the system:

- From deliberate disproportionate sampling of, for example, channel reception areas.
- In the case of **diaries**, to correct for an individual household's probability of selection, arising (for example) from random digit dialling whereby a household with two telephone lines would have double the probability of selection as a household with only one telephone line.
- Again for individual **diaries**, to correct for that individual respondent's probability of selection, for example where one person is selected in a three person household relative to a person living alone.

Weighting may also be required to correct for unplanned deficiencies in the structure of a sample, for example:

- Corrective weighting for differential non-response if this is established to be significant.
- To restore the sample's demographic composition, to conform to the latest published official estimates, for example in terms of:

- geographical structure
- sex balance
- age profile
- socio-economic or educational status profile

It is then possible to apply a **grossing-up factor** to the weighted data to arrive at population estimates, usually to the nearest thousand ('000).

Care should be taken to ensure that the range of weights arising from the procedures adopted does not become excessive, leading to instability in the data. Equally a weight applied to any household or individual should be relatively constant over time. Wildly fluctuating weights through time would indicate panel composition instability, making it unreliable for longitudinal analysis purposes. Details of weighting procedures and their rationale should be formally described and be fully transparent to users.

23. The Independent Audit

The independent audit is a critical element in validating the above range of quality control procedures as well as ensuring full disclosure of the methodology used by the research institute to users of the resulting data. The auditor should have access to all aspects of the production process, including individual data records, to assess how effectively and rigorously the quality control protocols are applied and the accuracy of research institute disclosures.

Independent audits should be required at regular intervals. The research institute then has a clear understanding that its procedures will be regularly checked.

24. Additional Data Collection – and Program Appreciation

Care should be taken not to impose on panelists an undue burden of additional data provision which could prejudice the quality of the basic audience measurement. Some broadcasters obtain a **program appreciation measure** within their audience measurement panel. It is important that such information be obtained in a non-intrusive way, and identically about the programs of all channels viewed in order to eliminate potential bias. This enhancement is an additional task requested of panel members. On the other hand it could motivate panelists to achieve high standards of compliance. It is recommended that the industry should continue to conduct experimental work to establish whether in practice the additional audience appreciation requirement affects viewing behaviour or push-button performance.

Push-button meter systems whose principal purpose is to measure TV audiences **should not gather product purchase information**. This stipulation does not rule out separate or supplementary single source panels whose raison d'être is to gather product purchase and media exposure data from a common sample. Another option is

data fusion of the audience measurement system on the one hand with the household product purchase or equivalent panel on the other.

Even in a one-week **diary** measurement system, the task of keeping the diary should be kept as simple as possible. The collection of demographic and other classification information should be limited to the strictly pertinent, in the interest of maximising response rates.

25. Viewing outside the Home

There is the issue of technically sound and cost-effective procedures for intermittent measurement of audiences to TV in places that are normally not covered by a peplemeter system:

- By the private household population in **non-domestic premises**, such as place of work, educational establishment, bar, hotel etc.
- By the private household population in **second or holiday homes**.

Intermittent measures of the above forms of incremental viewing are usually most accurately and most cost effectively obtained via studies separate from the peplemeter panel which employ:

- **Strict probability sampling** procedures for the sample selection ...
- a 7-day **self-completion diary** for data collection ...
- **Spread systematically over some 8-13 weeks** to represent the year as a whole (or 4 weeks to represent holiday home viewing in the peak summer holiday months) ...
- With a range of stratagems, including appropriate respondent incentives and personal placement and collection, to **maximise response**.

Recall is a realistic alternative, however, for broad measures of the audience to **specific TV events**, or regular series programming.

Whether the technique employed for a particular exercise is the diary or recall, it should be expected that some **calibration** of the data will be required to conform to the national peplemeter currency.

It should also be remembered that TV viewing taking place in a bar can be different in kind from TV watching in the comfort of one's own home. Its quality from the point of view of the advertiser is likely to be different accordingly. These are issues of professional judgement.

This judgement has, however, to be informed. It requires valid, empirical evidence of the residual audiences not measured by routine peplemeter systems. **Harmonization** in the detail of the data collection and reporting of individual surveys set up to provide

such supplementary data should therefore be a **priority issue** for all sectors of the industry: broadcasters, advertisers and advertising agencies alike.

The special technical challenges posed by viewing outside the home are addressed more fully by the ARM Group in their supplement to the second edition of the guidelines: “**Non-Domestic Viewing, Holiday Homes, International TV Channels**”, EBU (1997).

26. International and Niche Channels

Finally there may be the equivalent requirement in a market for intermittent measures of audiences to International and Niche TV Channels. They often do not have the penetration nor size of audiences to be satisfactorily measured by a country's basic peplemeter system.

The most effective route is again likely to be the high quality, probability sample **diary**. A fuller coverage of the rationale for the use of the diary technique for International and Niche channels will be found in the ARM Group's supplement to the second edition of the EBU guidelines referred to above (see VI.25).

It needs to be recognized that measurement of International and Niche Channel audiences is not easy. They are viewed by minorities – and by those minorities for (usually) only a minority of their viewing. It follows that registration of short viewing periods becomes a key issue.

Another technical challenge is effective coverage of the more mobile sectors of the population, if measurement of out-of-home viewing is a key requirement.

Potential understatement of their audiences, both in absolute terms and relative to other Channels, is therefore a major hazard for International and Niche Channels. Such technical difficulties need to be addressed directly in the choice of audience measurement technique to be employed, and in the minutiae of diary (or questionnaire) design (see VI.17).

VII DATA REPORTING

This section on data reporting has two purposes. One is to provide an ordered checklist of user requirements for specifying the outputs of television audience measurement. The other is to recommend standardized definitions of viewing measures and standardized formats of reporting. The adoption of standardized definitions and formats of reporting will assist the harmonization of national systems of measurement. The reporting conventions used in any particular market should be fully documented in its published Reference Manual for the system.

1. Universe Estimates

A combination of the national Census, some other high quality source, the Establishment Survey (if conducted) and the peplemeter panel itself are the basic source of data about the size and demographic composition of the population, channel penetration, means of signal reception and ownership of TV-related equipment. These data are vital for the use of subsequent records of viewing. It is recommended that these data are published or made available for electronic access according to the following criteria:

- Total universe sizes of households and individuals, including the publication of universe definitions.
- Population sizes in terms of each of the demographic classifications that are used in standard aggregated analyses.
- Penetration of all channels which are received by 1% or more of the population, irrespective of whether audience data are reported for them. In addition a complete list of all channels identified by the system should be included.

Channel Penetration. Percentage of TV households within the survey universe, or sub-universe, with set(s) tuned to and capable of obtaining a reasonable quality of reception of a particular channel, as quantified by the Establishment Survey or other sources. Channel penetration estimates therefore include illegal access to Pay TV, but exclude households with potential access to particular channels but whose TV sets are not tuned to receive them.

- Penetration (i.e. household ownership) of all TV- related equipment. It should be noted that penetration estimates can be affected materially by the household definition used, especially in under-developed countries. The list needs to include such items as number and location of sets in households distinguishing:
 - Portable versus static
 - Mains versus battery operated

- Colour versus black and white
- Stereo versus mono sound
- VCRs
- Teletext
- Interactive videotext
- Video games
- Video cameras
- Personal computers
- Internet subscription

The list is not exhaustive and will need to be extended following the introduction of new products such as widescreen television, high definition television, interactive compact video discs and PCs linked to the TV screen.

- Penetration of signal reception by cable (CATV), master antenna television (MATV), satellite master antenna television (SMATV), individual household satellite dishes (DTH), microwave distribution (MMDS).

Cable Television/CATV – Community Antenna Television. Wired transmission system in which TV sets are connected to a cable system serving multiple premises, and serviced by a cable operator.

MATV – Master Antenna Television. Wired transmission system in which TV sets are connected to a cable system serving multiple households in a single building (for example apartment block, or housing estate), which may or may not be serviced by a MATV operator.

SMATV – Satellite Master Antenna Television. Same as MATV, but with the addition of one or more satellite dishes feeding signals into the connected households.

DTH – Direct-to-Home. Reception by means of one or more privately owned satellite dishes, enabling households to receive satellite transmissions directly without the intermediary of cable transmission.

MMDS – Multipoint Microwave Distribution System. Over-the-air transmission (and reception) using microwave frequencies.

The distinctions between these different modes of reception are blurred at the boundaries. It is much easier to identify and differentiate between the various modes of reception in some markets than others.

Survey universes of cable and satellite households will sometimes employ somewhat different definitions in order to isolate certain specific kinds of viewing environment. For example, the universe of “cable” homes will sometimes include community antenna

(SMATV and MATV) systems. Or a single umbrella label, such as “satellite” homes, will cover all forms of reception in which viewers are able to receive satellite transmissions. Establishment data (or equivalent industry sources) should ideally report on each network separately, presenting statistics on channel penetration for each form of reception. Such data are important for understanding the structure of the survey reception universe.

In all but the most stable markets it is essential that penetration estimates of delivery methods (cable, satellite, etc.)—or any other technology which is known to affect viewing, and which is changing rapidly—be updated **frequently**. Monthly or quarterly reporting is a typical cycle.

2. A Viewer: Calculation Convention

The definition of a “viewer” of a particular channel or program at a particular time is determined by the TV set meter status and the push-button status, establishing respectively:

- Whether the TV set is switched on (or not).
- Whether a particular household member or guest is viewing TV at the time.

The peplemeter TV audience measurement technique in principle enables the audience to be measured second by second. However it should be recognised that, irrespective of finer issues relating to metering and processing technologies, second by second data rarely have the precision implied by the term. While it is useful to have access to such data for certain specific ad hoc analyses, it is neither necessary nor appropriate for standard data processing:

- The commercial significance of establishing viewing down to the individual second is questionable, nor is it at all necessary for programming purposes.
- The synchronization of the TV set meter clock will not always be perfect to the nearest second.
- Program and commercial break start and end times will often not be accurate to the nearest second.
- The panel member or guest will not register precisely when he/she begins and stops viewing television to the nearest second. There will usually be lags in push-button status registration.
- Routine data processing at the individual second level is a relatively expensive option.

For these reasons, virtually all peplemeter systems use averaging and simplifying arithmetical conventions in processing audience data. In-panel coincidentals, together with observational studies conducted in peplemeter households, are invaluable in providing empirical guidance towards the development of appropriate bridging and crediting rules for these conventions. Current practices vary greatly from country to country, for example:

- Exact cumulating of logged second.
- Exact cumulating of logged seconds for the most recent logged change at the end of each set time unit (for example 30 seconds) by counting back to the time of the change, the remainder being assigned to the previously recorded set status.
- Cumulating of “snapshot” readings of set use (for example every 5, 15, 30 or 60 seconds) and attribution of the surrounding interval to the channel registered by the snapshot.
- Cumulating of assigned time units (for example a clock minute), each being determined by the majority attribution of set status by periodic polling (for example every second or every few seconds).
- Cumulating of assigned time units (for example a clock minute), each determined by the last registered set status during that unit.

There are two elements to the conventions:

- The persistence thresholds for registering a change in set status and individual presence as viewer.
- The algorithm (i.e. arithmetical procedure) for calculating audiences.

It is unlikely that the variations in the conventions for calculating viewership are the cause of significant differences between different systems. Overall differences in audience size are likely to be very small, since differences only arise at the beginning and end of viewing sessions. However, they could slightly affect the balance of viewing across channels, especially during commercial breaks.

The calculation convention used should always be fully documented in the Reference Manual.

3. **Components of Viewing**

In setting out to measure TV audiences, it is important to distinguish:

- Each identified channel whose live audiences are reported routinely and individually
- The sum of identified channels whose live audiences are too small to be reported individually, sub-divided into:
 - Other terrestrial
 - Other cable/DTH
- Timeshift viewing – which may or may not be aggregated with live viewing to form “consolidated viewing” (see below)

- Total TV viewing – live or consolidated
- Each identified other use of the TV set, for example teletext and TV games

- Total identified uses of the TV set
- Total unidentified uses of the TV set

- The total TV usage

This conceptual framework is necessary to obtain trans-frontier consistency in reporting TV audiences, including the total TV audience.

To this end it is necessary to:

- Include viewing for all sets in a household, which are capable of being metered (see VI.12 above).
- Include guest viewing (see VI.15).
- Include all viewing to a channel where tuning has occurred via the VCR.
- Where the decision has been taken to include timeshift viewing of channels, show separately “live” and “consolidated” audience estimates. In deciding whether to include timeshift in the published viewing figures, systems need to consider the trade-off between the greater accuracy of the eventual audience estimate versus the extra delay in publication caused by the need to allow extra time for timeshift viewing to occur.
- Retain panel members who are absent or away on holiday within the in-tab samples.

Recommended definitions for live, timeshift and consolidated viewing are as follows:

Live Viewing. Viewing of a broadcast at the actual time of transmission. This applies to any form of reception, including off-air, cable, community antenna or satellite dish, and use of the VCR as tuner.

Timeshift Viewing. Viewing of a television transmission recorded on video cassette after the actual time of transmission (i.e. the playback of a recorded TV program(s)).

Consolidated Viewing. Live viewing plus timeshift viewing (i.e. playback of video recording) within a set time interval from the original transmission. Consolidated viewing may involve double-counting. A panel member may view a program more than once (for example both live at the time of transmission and timeshifted, or timeshifted twice).

In the event of a system reporting consolidated viewing, the decision on precisely how many days to allow for timeshift viewing to occur is necessarily arbitrary. As noted above, the more days allowed, the more complete is the viewing figure at the expense of speed of data publication. The state of current empirical knowledge suggests 7 days as the best compromise.

4. **Audience Measures**

A TV audience measurement system sets out to measure the **reality** of TV viewing. However, such reality is highly complex. Over the years, therefore, users both for editorial and for commercial purposes have developed a number of **conventional measures** of TV viewing. Out of these conventions have developed specific **terms**. It is the purpose of this section to list and define the key audience measures that are required both by the broadcaster editorial community and for the trading of commercial airtime. A number of these measures can be expressed both as a **percentage** of a population and as a **projected population estimate**:

- Ratings (average audience), GRPs (gross rating points) – and HUTs/PUTs
- Audience composition in terms of demographics
- Cover/Cume (cumulative audience)/Reach
- Frequency
- Amount of viewing
- Channel Reach
- Audience share
- In addition, many television stations (especially public service broadcasters receiving public funding such as a licence fee and with the remit to cater for all the public) require a measure of substantive viewership.

The following definitions are recommended:

Rating. The average percentage of a population viewing (or projected population estimate), as estimated from the sample, viewing across a defined time period.

GRPs – Gross Rating Points. The sum of the individual spot ratings in an advertising campaign. Also known as ratings, rating points, TRPs (Target Rating Points) and TVRs (TV Ratings).

HUTs/PUTs – Households/People Using Television. The average percentage of households/people using television (in total across all channels) across a specified time period.

In this and the following definitions, “population” can refer to the total universe population of all individuals or any specified sub-population/target audience.

Most peplemeter systems deploy arithmetical procedures (see VII.2 above) that compute Ratings minute-by-minute. The Rating for a specified time period will be the average of the minutes belonging to it.

Ratings can relate to the total TV audience, averaged over a specified time period expressed as a percentage of the population, or for example to a specific program within a particular channel. Some countries have a special term for the total Rating – separately for household data and individuals’ data:

- HUTs – % of Households Using Television
- PUTs – % of People Using Television

A further set of operational definitions is employed for assigning Ratings to advertising spots. Options include crediting the spot with:

- The Rating of the minute in which it starts
- A weighted average of the minutes in which any part of it occurs
- An average of the exact seconds which it occupies
- The average of the whole break in which it occurs.

These different procedures lead to very different individual spot ratings, and can lead to systematic differences even when spots are considered in aggregate, such as for a complete campaign. The precise technique used in each country is usually decided after discussion between advertisers, agencies, broadcasters and the research company, and will be determined by national requirements. The best solution is that which provides the closest measurement of viewing during the precise duration of the spot, consistent with providing data which can be handled robustly by the analysis software.

Cover/Cume/Reach. The cumulative percentage of a population (or projected population estimate) registered as viewers to a sequence of programs or schedule of advertising spots at least once.

Cover, Cume and Reach are identical concepts. The terms are interchangeable.

Fundamental to the concept of Cover/Cume/Reach is any viewing. In other words, anyone who contributes to the Ratings of a schedule of advertising spots or programs will, by definition, belong to its cumulative audience.

The analysis of program audiences may require qualified definitions of Reach. For example, producers and schedulers may be less interested in pure Reach/Cover, where any viewing counts, than in the numbers and demographics of those who watch at least (say) 15 minutes of the average 30 minute program, or 50% of its duration.

Depending on what is wanted, variable time thresholds may be set for determining program or channel Reach. The thresholds may be absolute (minutes and seconds) or relative (percentage of transmission). In addition, users may want to differentiate between non-consecutive and consecutive viewing when defining their reach criteria. The setting of thresholds is necessarily subjective. It is recommended that, where variable thresholds are used, they are specified in the labelling of analyses (e.g. "five minute program Reach" etc.).

Frequency. The average number of times that members of a population, who have been counted at least once as viewers to a schedule of advertising spots or sequence of programs, have counted as viewers.

For advertising purposes, the terms "Reach" and "Frequency" have a distinct and specific meaning.

Reach is the percentage of the target population reached by a campaign or schedule (the words are usually interchangeable), that is, the proportion of the target population who had the opportunity to see at least one of the commercial spots in the campaign. This basic measure is often called "1+ Reach".

Other levels of Reach are also commonly used in measuring advertising schedules. These measures are often known as "Effective Reach", and can be expressed as "n+ reach", or "n-m reach". As such, they combine measures of reach and frequency (with the implication that one exposure is insufficient to fulfil the objectives of the campaign). N+ reach is the percentage of the target population who have seen at least n of the spots in the campaign. N-m reach is the proportion of the target population who have seen at least n of the spots, but not more than m.

Frequency, in advertising terms, is therefore the average number of spots in a schedule which each member of the target audience sees (Gross Rating Points divided by net 1+ reach). This is often expressed as average frequency, or (average) OTS (opportunities to see), but frequency can also be applied to any level of exposure. For example, 4+ cover at 8 OTS means the number of people reached at least 4 times, with an average frequency of 8 exposures each. Gross OTS, or gross frequency, is the same as total GRPs.

The key measurements for any advertising campaign are:

- Net reach, at 1+, 2+, etc levels
- Total GRPs
- Average frequency or OTS

The provision of these simple measures is more difficult than it appears. This is due to the problem of missing data.

Each day a small number of viewing records will not be retrieved for one reason or another: perhaps 2-3%. A user may want to carry out a (say) six week Reach and Frequency analysis. The question is: what should be done about the small number of viewing records missing each day that cumulatively may reduce the sample size available by as much as a third or a half?

One solution is to restrict the analysis to what are called “continuous reporters”: those for whom a complete data set exists. If this is done, it is likely to lead to significant bias in the results. Missing information is more likely to arise from larger and more complex households, and they will probably be under-represented amongst the continuous reporters. This problem becomes progressively worse, the longer the period being analysed.

An alternative and technically superior solution is to impute data and to assume nil viewing for every record that is missing. The effect of this procedure on schedule Reach at the 1+ level is likely to be minimal. Very few spots in a campaign have a high unique cover: that is, the only spots in a campaign which a member of the target audience sees. There will, however, be a slight understatement of frequency. This can be adjusted by modelling, if required. It may also be necessary to set a limit on the degree of missing data which is allowed before a respondent is removed from the database.

Whichever of these procedures is followed, the spot ratings generated by the reach and frequency analysis will never match those produced by the audience measurement system itself, because they will be based on different samples. If published ratings include guest viewing, there will be an even greater discrepancy (because guests never form part of the data set used for reach and frequency analysis). It is therefore also necessary to decide whether there should be further modelling to make sure that

they do match. This is not a technical decision. Some countries like to make sure that all spot ratings are always consistent, while others are content to display the inevitable inconsistencies.

A further complication occurs if, as is common, the viewing data are weighted (to correct for uneven response) or grossed up to universe levels. Each individual in the sample will usually be given a unique weight each day. For Reach and Frequency analyses, each individual requires a single weight to be applied to all days of the analysis. A simple and useful convention is to take the weight for each individual at the mid-point of the span of the analysis. If this is done, then only those respondents present on this day (regardless of whether or not they actually watched any television on that day) will be included in the analysis.

In each country and market the problem of missing data, and how precisely Reach and Frequency should be calculated, needs full debate and a consensus reached amongst users from all sectors.

Amount of Viewing. The amount of viewing (weekly) per head of population, by channel or in total, and expressed decimally in hours: live or consolidated.

It is recommended that total viewing refers to television sources only, and excludes all other uses of the TV set: VCR use (except timeshift, where consolidated ratings are reported), video games, teletext etc. The week is recommended as the basic time interval for reporting, though this may be subdivided (for example weekday versus weekend).

Audience Share. The amount of viewing obtained by a particular channel or program as a percentage of total viewing by a population across a defined time period.

Again, it is recommended that total viewing should refer to television sources only, and exclude all other uses of the TV set.

Channel Reach. The percentage of a population (or projected population estimate) viewing a television channel across a specified time period for a specified minimum number of minutes.

This is in essence a measure of the number of viewers that a channel has. The precise definition employed will need to depend on the purpose for which the measure is to be used. It will have two components:

- The **period of time** covered: for example, a week or a day.
- The **number of minutes** of viewing to qualify within that period: for example, 5 minutes or 15 minutes.

An additional technical issue arises out of the second of these elements: should the minimum number of minutes required be continuous or not? In practice it has been established that there is rarely much material difference between the two. It is probably simpler, therefore, to set the requirement as a **total** (not continuous) number of minutes.

The standard definition of Channel Reach requires a **minimum of 5 minutes viewing of the channel in total** – across the day or week, as appropriate. The 5 minute Weekly Reach figure for a channel represents its potential audience available to the advertiser.

For broadcasters receiving **public funding** (for example, a licence fee), Channel Reach is a key accountability measure. A more substantive definition will be required. A minimum of 15 minutes may be considered appropriate: equivalent to at least half of a single 30 minute program. **A 15 minute Weekly Reach** represents, therefore, viewing of at least one of its programs across the week.

Much shorter definitions can be misleading and should be avoided. For example, a one minute Reach will usually require no more than 15 seconds of tuning to the channel across the day/week (see VII.2). It will be including a degree of casual channel zapping. However, in some markets (e.g. in North America) it is the custom to require panelists only to register when they regard themselves as viewing a program. Under these circumstances shorter definitions of Channel Reach do not pose the same problems.

5. Availability of Data for Analysis

The organization responsible for data collection should ideally, in addition to providing data processing services itself, make data available to third-party bureaux.

In JIC and MOC systems (see V) it is in principle in the interest of users that contractual arrangements in any country or market permit and encourage the development of a wide range of analysis systems available to authorised users, where they have the right to choose the software system they wish to use. Competition is an effective means to this end. It is helpful, therefore, if contractual arrangements enable **competitive licensed data bureaux** to be set up, giving scope and opportunity for initiative and creativity to flourish in the development of software tools. The more value is derived from an audience measurement service, the more users will be prepared to invest in it.

However, such arrangements then require certain disciplines. A condition of such a third-party licence should require a formal commitment to whatever are agreed to be the **“gold standard” algorithms** used in the calculation of specific audience measures such as those listed above (see VII.4). A Data Processing working party, representing all interests, is a useful means to this end.

Options for making audience data available include printed reports and computer files. They need to include four distinct elements:

- A viewing records file.
- A program log: the names, start and end times and genre (see VII.10) for each program transmitted.
- A commercial log: the equivalent for commercials.
- A household master file

Electronic reporting can be supplied by means of disk, CD-ROM, on-line electronic access and down-loading, or other forms of electronic transfer. In most countries a growing number of software packages is available for local PC use and analysis.

The choice of output system will depend on individual market needs for detail as well as on general issues of cost. The rapidly falling cost of computer power is however becoming less and less a constraint for effective use of a system. Users need readily accessible **standard analyses** (i.e. available to all licensed users in standard demographic breaks and in a range of reporting formats), and access to other data via **special analysis or on special request**.

For standard analyses, users need to decide:

- Which channels and universes are to be reported.
- Which demographic and other individual variables are to be chosen for analysis.
- Which time periods are to be selected for presenting the results.

Peoplemeter systems vary widely over the number of standard demographic (and other individual) breaks that they offer for analysis. The majority offer from 10 to 60 standard reporting breaks per country. In any particular market the industry needs to reach a consensus on the precise demographic breaks and their number that users need as a standard service – whether to meet editorial or commercial needs, taking account of the cost implications of these requirements.

With the emphasis shifting from printed to electronic reporting and PC-based packages, requirements are tending more and more towards fully flexible systems so that users can choose their own, demographic customized break-outs. Standard reports will nonetheless be needed for some years to come as convenient, economical summaries of national data.

Aggregated Data. Pooled individual viewing statements that have been summed and weighted in order to produce audience estimates for specified populations/target groups.

Aggregated data provide the analyses of ratings, audience share, amount of viewing and other variables of audience size. Whereas cover and frequency calculations from

raw data are longitudinal, aggregated analyses are cross-sectional. For many (standard) purposes it is economic to store aggregate data. No information is held on individuals.

Disaggregated Data. Respondent-level data after editing and validation.

Users must have access, by some route, to respondent-level data. Disaggregated data in peplemeter systems contain four kinds of information:

- Record of set status for each measured set in the household.
- Presence of individual as viewer in relation to each set.
- Demographic and other information about the individual's identity.
- The individual's daily weight.

Disaggregated data are commonly used for reach and frequency analysis, where it is important to know the degree of duplicated viewing across two or more programs, commercials, or time periods. Similarly program and channel measures of reach require access to disaggregated data. (The term "raw data" should be applied to the unprocessed household set and individual meter records as received by the central processing unit – prior to quality control procedures.)

Conditions of access will depend on the organization of the system and the contractual arrangements between the customer and data provider (see V above). Within the basic contract price for JIC/MOC systems, reporting arrangements will need to include identification and start/end times of both programs and commercials – on a non-discriminatory basis to all users. Beyond this it is essential that access is provided at a reasonable cost – including to users abroad. It is recommended that for JICs and MOCs the amounts charged cover the pure cost and a reasonable profit for the data provider. Such costs should be able to be reviewed by an independent body representing all users.

6. Speed of Access and Reporting

The speed of access to and reporting of data is an important element of the contractual arrangements between the customer and data supplier – including issues of copyright, ownership of the database and the right of either party to undertake further analyses.

Programming and advertising sectors require access to data at the earliest opportunity. Most systems are able to report audiences for time periods, commercials and programs the next day by means of electronic access. This should be the aim of all systems. However, definitive Ratings may take longer to produce, for the following reasons:

- Data suppliers will not succeed with the overnight call-up procedures for 100% of panel homes. A few will have to be tackled again the following night. As a result the overnight (i.e. next day) Ratings will not be definitive.

- Commercial and program Ratings require transmission logs (or equivalent sources) for matching viewing records against the actual transmission times of programs and commercials. This may take longer than the reporting of time period Ratings, depending on market complexity and the depth of detail that is wanted.
- Where the system publishes consolidated Ratings, an extra period has to be allowed for timeshift viewing to take place. The recommended interval for timeshift viewing is 7 days (see VII.3 above).

The evolution of the television medium is likely to increase the opportunity to view on demand or to view same-week repeats. Programs that are available to viewers at their discretion will then accumulate audiences over a protracted period of time. Individual time-based ratings may then become less relevant than some form of (net) weekly reach under this scenario.

7. Reporting of Channels

Cost considerations and terms of payment will affect the reporting arrangements for individual channels. However, it is also important to recognize that national markets are rarely able to sustain more than one audience measurement panel. Therefore it is important that systems accommodate the needs of as many channels as possible.

It is recommended that systems should report basic **reach and share figures by principal demographics for all channels available in any market, and viewing data in full for all television channels that achieve at least a 5% Weekly Reach whether or not they subscribe to the service** – but subject to sample size limitations (see VII.11).

The industry in a particular market needs to determine by consensus precise rules for their own market, and in particular to agree to equivalent thresholds for reporting of niche channels. For example, it may be appropriate to report individual Children's Channels that achieve a 10% Weekly Reach amongst children – but again subject to sample size considerations.

8. Universes for Reporting

Most systems report Ratings across more than one Universe. Except where panels are not national, all systems report audiences against their national universe, but may also report audiences for a number of sub-universes, for example:

- By geographic region.
- By populations speaking different languages.
- By satellite or cable households.
- By individual satellite/cable coverage areas.

The selection of universes for reporting will significantly affect the Ratings. For example, a satellite channel received by 50% of TV homes nationally would obtain double the reported Ratings against this universe compared with the national universe. It is recommended that systems publish channel penetration figures separately against each universe that is used for reporting viewing data.

9. **Program Classification**

An important requirement for senior management in broadcaster organizations and their schedulers is a program classification system. The audience measurement database needs to hold a program genre code for each program transmitted each day all 365 days of the year – for all significant channels in the country or market. The logistics for achieving this will need to be agreed between the broadcasters to achieve consistency. At minimum the following macro-genres need to be coded:

- News
- Factual – including current affairs and arts
- Education
- Films/movies
- Other Fiction – including drama, soaps, comedy
- Entertainment – including talkshows, gameshows, music
- Sport
- Children's programs
- (Commercial Advertising)

It is also useful to code all programs across all significant channels by country of origin:

- Own country
- Elsewhere in Region (for example, Europe)
- USA
- Elsewhere

10. **Demographic Variables for Analysis**

Most audience measurement systems offer any combination of coded variables for conducting special analyses, subject to the sample sizes available (see VII.11) in addition to the selected codes for standard analysis.

In addition to whatever reporting conventions are felt to be appropriate and are the tradition within the country in question, for between-market comparative purposes it is recommended that the coded variables should include:

- All individuals – aged 4 and over
- Children – aged 4–14
- All adults aged 15 and over (Men, Women, Chief Income Earner, Housewife/Housekeeper/Main Shopper/Lady of the House)

- Age
- One or more sociodemographic scales
- Household size
- Ethnic classifications, where significant
- Number of television sets
- Region
- Conurbation, urban, rural
- Household reception, where relevant (for example cable, DTH, etc.)
- VCR homes (and other types as are appropriate to the country in question)
- The geographical universe for each Channel

Note that it is recommended that:

- Data should be reported or be available for all individuals aged 4+.
- For trans-frontier purposes a threshold of 15+ be set for adult viewing. Arguments can be made for other thresholds. This recommendation is simply in line with most common practice.
- Standard age-breaks beyond 15+ should at minimum be by ten-year intervals (viz. 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+), with other groupings to meet individual market requirements but subject to sample size constraints. In reporting audiences by age and other demographic breaks in standard analyses, care should be taken to ensure that the chosen breaks correspond with the categories by which the panel is balanced. Otherwise there is an increased risk that the reported results will not correctly reflect the survey universe. Surveys should code exact age so as to permit fully flexible age specification in special analyses.

The industry should on an international basis work towards developing common and standardized definitions of key demographic terms: for example “household” and “housewife/housekeeper/main shopper/lady of the house”.

Main Shopper. The individual person, either man or woman aged 15+, who is most responsible in the household for purchasing household provisions.

- National systems should adopt wherever possible standard international sociodemographic criteria for classifying individuals, in addition to their own national sociodemographic classifications where different. The priority of national surveys must necessarily be to satisfy the requirements of the domestic marketplace. However, the goals of harmonization require the development of classifications under (for example) ARF/ESOMAR auspices.

11. Small Samples.

Recommendations on minimum sample sizes for peoplemeter systems have been made in VI. 9. Decisions on where to set the boundaries for reporting of data are ultimately subjective. So long as the limitations of small samples are recognized, small samples can yield useful insights. Consequently, it is recommended that no absolute limits be set on sample sizes for reporting. However, it is recommended that data based on a sample size of less than 75 individuals be flagged, and that none be used for commercial trading purposes below a threshold of 50 individuals.

Estimates of the confidence limits that apply to a television audience measurement system should be published – ideally in an annual technical reference manual.

12. Time Units for Analysis

Standardized time units are needed for reporting of audiences and for calendar periods.

Again local markets will have their own appropriate time unit conventions. **In addition** standardized units of reporting in standard reports and/or special analyses need to include:

- minute-by-minute (in the case of peoplemeter systems).
- quarter hour.
- daypart: 0300-0600, 0600-0900, 0900-1200, 1200-1500, 1500-1800, 1800-2100, 2100-2230, 2230–0000, 0000-0300.
- commercial break (for peoplemeters).
- commercial minute (for peoplemeters).

The following are recommended as standard time definitions for reporting audience data:

Day: 03.00 – 02.59 hours
 Week: Monday to Sunday, the main subdivisions being:

- Monday to Friday
- Saturday
- Sunday

in addition to the weekdays individually.

Month: 4/5 weeks (12 months per annum) on a systematic cycle:

for example	March	As 5-week months
	June	
	September	
	December	

and the remainder as 4-week months.

A somewhat less satisfactory but acceptable practice used in a number of countries at the moment is the 4-week definition. The disadvantage of this convention is the lack of linkage then between a 4-week month and the calendar month, and the inconvenience of the year then comprising 13 months indivisible by 4 to yield quarters.

Year: 12 months (a week 53 every 7th year) with week 1 of month 1 (by definition) beginning on a Monday within a range of 28 December to 4 January, standardized for adoption by all countries.

VIII THE FUTURE

Preparation of the earlier editions of the EBU guidelines on audience measurement was prompted by several considerations. The needs of many countries had become more complex, more competitive, and consequently more demanding on research methods. Peplemeters provided an answer, but were comparatively new and untried. Meanwhile multinational market trends called for greater comparability of national market data.

The guidelines aimed to recommend standardized and optimal procedures where possible for measuring audiences, and to bring into the open methodological issues underpinning peplemeter measurement. **It was also intended that they should be reviewed in the light of subsequent experience and market developments.**

The aims of this third edition of the guidelines have stayed the same but are explicitly global in perspective. They too will need to be reviewed and to evolve as experience builds and television itself continues to evolve.

In looking to the future, two issues stand out: first, the prospects of further radical transformation of television viewing conditions, and secondly, the capacity for development of the research tools for measuring audiences.

Up to now, the chief causes of greater complexity and competition have been satellite and cable television. They are largely responsible for the widespread expansion and segmentation of channel choice, the loosening of advertising restrictions, and the emergence of subscription-based services. Satellite and cable had symbolized the first cycle of transformation which has still some distance to run in many countries. A second cycle is heralded by the advent of digital compression, multiplexing and convergence of television and computer technologies. These developments promise to create the potential for hundreds of channels and trigger a major growth in the number and variety of interactive services.

Today, it is still customary to treat television largely as a medium that people watch at home. The over-riding function of audience measurement has been to estimate mass audiences for channels, programs or commercials. It soon may be necessary to establish how people use the television set outside as well as inside the home, in addition to measuring mass audiences for traditional telecast services. The television medium may emerge from all of this in a dramatically different form.

It is hard to predict how this second cycle of transformation will shape future research requirements. Two implications at least appear certain.

First, it will increase the complexity and variety of viewing conditions in the home, with the need for conventions to be developed on the reporting of time spent zapping between channels. This will both create new challenges and impose a significantly greater strain on sampling procedures to meet the existing demands for measuring mass audiences. Either way, an upward pressure on research costs can be expected. Fortunately, however, we can expect the cost of data processing to continue to fall.

Second, digital compression will cause a need for audience research companies to make sizeable investments in meter design to ensure that current peplemeter systems can adapt to digital compression and transmission. The systems will have to be provided with new techniques that will ensure correct channel and program identification, as the direct relationship between measured voltage and channel will not necessarily exist. Indeed, although new techniques such as passive meters may offer the possibility of major advances in audience measurement, it is important not to lose sight of the fact that peplemeters, which are today's industry standard, also face serious challenges. Without a major investment in research and development, electronic measurement may not be able to contend with the technical changes brought about by the introduction of digital TV. Similar practical considerations apply to other research developments, such as supplementary studies of out-of-home viewing, data fusion and single source data.

Just as in the past, so in the future television audience measurement must be based on sound scientific theory. The Ten Principles enumerated in section IV of these guidelines must continue to be observed scrupulously by the industry – and by data suppliers in particular.

Common to all these endeavours will continue to be the ambition to generate data on viewers that is more diagnostic for broadcasting, programming and/or advertising purposes than what is presently available. How far and in what direction the new methods proceed will depend on the evolution of markets and market needs, and on the critical issues of cost and the industry's research budgets. One of the main problems of the future will certainly be the growing complexity of research requirements. Solutions may well exceed budgetary resources. The length and duration of national audience measurement contracts will in the meantime need to take account of such issues.

However there are opportunities for positive developments too.

The number of "players" amongst whom costs can be shared will continue to grow apace. There will be the explosion in the number of TV channels available to the public emerging out of the digital revolution, whose audiences need to be measured – whether funded by advertising, subscription or pay-per-view. Equally the number and range of suppliers of goods and services seeking TV advertising will continue to grow.

There is also opportunity for the “players” themselves to contribute a constructive role in easing the industry’s audience measurement task. Broadcasters can assist in the identification of their channels and programs. Television manufacturers in the development of the digital TV set and set-top boxes can take account of the engineering implications for audience measurement. Enlightened self-interest has a part to play, arising from a better understanding of the industry’s common interest in measuring TV audiences.

The ARM Group hopes it will have contributed to these endeavours.

The Audience Research Methods (ARM) Group**Advertising Research Foundation (ARF)**

641 Lexington Avenue
New York
NY 10022
USA

Tel: +1 212 751 5656
Fax: +1 212 319 5265
E-mail: www.arfsite.org

President: Jim Spaeth

Canadian Advertising Research Foundation (CARF)

175 Bloor Street West
South Tower
Suite 307
Toronto
Ontario M4W 3R8
Canada

Tel: +1 416 964 3832
Fax: +1 416 964 7645
E-mail: www.carf.ca

Chairman: Geoffrey Parker

European Association of Advertising Agencies (EAAA)

5 rue Saint-Quentin
B-1000 Brussels
Belgium

Tel: +32 2 280 1603
Fax: +32 2 230 0966
E-mail: info@eaaa.be

Director General: Stig G. Carlson

European Broadcasting Union (EBU)

EBU
Ancienne Route 17A
CH-1218 Grand Saconnex GE
Switzerland

Tel: +41 22 717 2007
Fax: +41 22 7474000
E-mail: ebu@ebu.ch

Secretary General: Jean-Bernard Münch

European Group of Television Advertising (EGTA)

Rue Wiertz 50
B-1050 Brussels
Belgium

Tel: +32 2 286 9122
Fax: +32 2 286 9124
E-mail: yolande.bogaerts@egta.com

Secretary General: Michel Gregoire

European Media Research Organizations (EMRO)

Secretariat:
Oberdorfstrasse 28
CH-8114 Dänikon
Switzerland

Tel: +41 1844 0809
Fax: +41 1844 2176
E-mail: herbertfurrer@bluewin.ch

Chairman: Jürgen Wiegard, AG.MA, Germany
Secretary: Herbert Furrer

European Society for Opinion & Marketing Research (ESOMAR)

J.J. Viottastraat 29
NL-1071 JP Amsterdam
The Netherlands

Tel: +31 20 664 2141
Fax: +31 20 664 2922
E-mail: email@esomar.nl

Director General: Jürgen Schworer

Group of European Audience Researchers (GEAR)

Secretariat:

Ancienne Route 17A
CH-1218 Grand Saconnex
Geneva
Switzerland

Tel: +41 22 717 2007
Fax: +41 22 747 4000
E-mail: ebu@ebu.ch

Chairman: Wim Bekkers, NOS, The Netherlands
Secretary: Paolo Baldi, EBU, Switzerland

Pan-European Television Research Group (PETV)

Eurosport
55 Drury Lane
London WC2B 5SQ
England

Tel: +44 171 468 7777
Fax: +44 171 468 0023
E-mail: rshaw@eurosport.co.uk

Chairman: Rachel Shaw

World Federation of Advertisers (WFA)

Avenue Louise 120 – Box 6
B-1050 Brussels
Belgium

Tel: +32 2 502 5740
Fax: +32 2 502 5666
E-mail: info@wfanet.org

Director General: Stephan Loerke