

LESS QUESTIONS, MORE DATA: *REVITALIZING THE EUROPEAN CURRENCY IN SINGLE SOURCE AFFLUENT AUDIENCE MEASUREMENT*

Harm Hartman, Synovate

Abstract

Audience research has experienced a growth in the number of media types and titles over the past decades, along with a drop in response rates due to a number of factors including long and boring questionnaires. EMS, the European currency for the affluent population, developed a mixed mode approach to break this downward trend by asking only 50% of all media questions. A complete data set is achieved by data fusion based on Chaid trees. Every title is modeled by its discriminating variables to reproduce all levels of and relations between titles and relevant target groups. Larger sample sizes are needed when fewer questions are asked, but shorter questionnaires provide higher quality data: we ask less and get more.

Keywords

Data fusion, Chaid, audience research, mixed mode, weighting, EMS.

1. Introduction

While response rates are dropping the need for single source information is still growing: we want answers to more and more questions. Our primary source of information, the respondent, is however less and less willing to participate in long and boring questionnaires. This is especially true for audience research where the number of media titles for Print, TV and Websites has increased dramatically and questions on reach are considered to be boring and repetitive. For EMS, the widely accepted currency in European elite audience measurement for Print, TV and Web, the solution for this problem was resolved by increasing sample sizes (partly online), asking less questions and using data fusion techniques to complete the data sets.

2. Background

2.1 History and future of data collection

Looking back on the past half century of market research we see that technological innovation has been the driving force for major changes in the way we collect data. As soon as large main frames computers became available for data entry and analysis, market research agencies used them for analysis. After that information technology continued to become more accessible, such as the (still very large) mini's, the desktop computer, the laptop, the internet and now smart phones and registration techniques (log files) at every stage the market research industry used the newest technologies for data collection (table 1).

Decade	Innovations in data collection
'60-s	main frames: data-entry, tabulation, multivariate analyses
'70-s	mini's: computer assisted interviewing (CATI)
'80-s	desktop's: large growth in market research agencies offering CATI
'90-s	laptop's: computer assisted personal interviewing
'00-s	web interviewing / online access panels
'10-s	smart phones / registration techniques: fusion?

Table 1. decades of data collection

The introduction of these innovations in data collection were mainly cost driven, but had significant effects on the quality as well.

- Face-to-face using paper-and-pencil was very cost effective when salaries for interviewers were low (until 60's)
- Postal / paper was cost effective when response rates were relatively high.
- When telephone interviewing became popular it was very cost effective compared to face-to-face interviewing at home.
- With random digit dialling (RDD) access to all households with a fixed telephone-line was achieved.
- Computer-assisted interviewing had a positive impact on the quality of the data: automatic routing, input control, consistency checks, immediate processing of the data.

With the growth of the market research industry, the number and length of interviews conducted also increased, which in turn increased the respondent burden. Non-response became a serious threat for data quality, not only because of possible higher selectiveness when the task of the respondent is too heavy, but also because higher drop out means higher costs, more weighting and less reliable results.

In the 00's, online interviewing appeared to be most cost effective with mainly fixed costs and little variable costs. However, a good online alternative for probability sampling was hard to find. The solution of creating random samples was found in panel research (online access pools), but this requires loyalty programs: respondents are paid to cooperate in surveys. The professional respondent was born and serious doubts on the quality of online panels rose (Nopvol). Mixed mode data collection creates possibilities for adjustments of mode effects (other respondents, other answering behaviour), but can be rather costly (with additional fixed costs and extra handling with different fieldwork management programs).

Nowadays, in the second decade of the 21th century, we still see new techniques entering the arena of data collection: smart phones, gaming exercises, and all kinds of (online) registration techniques. The question is, will this decade be seen as yet another decade of data collection or as the decade where the problem of **'how to handle new and fragmented data sources!'** was fixed.

Survey research has experience large innovations in primary data collection, which has been responsible for a large increase in the number of respondents and in-directly for the high non-response causing serious doubts on the reliability (stableness) and validity (selectiveness) of (panel) research results. Though it seems that more data is available, it is based on selective and fragmented sources. In the near future methods are to be deployed dealing with selective samples and registration data, such as propensity scaling techniques and other weighting/imputation methods. The question we are dealing with now is **'how can we still be able to distillate valid and reliable research if respondents are not willing to share all information at once?'** Smart survey design and intelligent data fusion are solutions to cope with this.

2.2 Nature of audience research

Information on media behaviour for relevant target groups is the base for media planning and single source multi-media information is still the ultimate solution.

Traditional Audience research has the objective to create an indisputable and affordable currency for the media industry in terms of opportunity to see:

- Reach and frequency of all media titles to obtain probabilities to see (read Print, watch TV, visit Websites)
- Frequent releases to show (stable) trends by smoothing wave data
- Large sample sizes to report also small media titles on relevant target groups

Buyers of audience research, the media owners and agency planners, are looking for single source information at a reasonable costs. Acceptance is based on:

- Representativeness: correct distribution of target groups plus the possibility of projecting to actual population numbers
- Reliability: stable results within an acceptable margin of error
- Validity: in accordance with real media behaviour
- Accessibility: possibility to do re-contacts

To be accepted by the market implies a transparent and solid methodology of data collection and data processing at an affordable price. The paradox of audience research is that on the one hand a currency has no uncertainty or standard reliability intervals, but it is accepted as a sort of reality. On the other hand the industry accepts that the practice of research can't be perfect, and has a bias, as long as it is stable over time.

The method of data collection is a trade off between accuracy/reliability on the one hand and costs/investment on the other hand. In the past decades costs have driven the audience standard from face-to-face interviewing into telephone interviewing, which allowed for larger sample sizes but forced the interviewer to read out long lists of media titles. In the UK and US we see that face-to-face interviewing (CAPI) is still the standard for audience research, but around the world CATI has been accepted as an affordable alternative to CAPI. Experiments with online (panel) surveys for audience research still suffers from lack of credibility, due to questionable or unknown panel quality, plus a perceived unfair advantage given to online media over traditional print media and possible method effects. Experiments with online audience research had mixed successⁱⁱ. In the Netherland the National Print Monitor (NOM) which was almost solely based on online panels has been stopped. The new solution is found in combining fresh sample (CAPI and online) with additional panel research, a combination of cost effectiveness but within standard probability sampling frame.

Audience research based on passive capture (registration of television- and internet behaviour) has been in the market for some time, but especially for Print and multimedia surveys this is not always feasible. The current mainstream methodology for audience research is in its nature unattractive for respondents: for reach and frequency all titles have to be asked, and the number of titles (and media platforms) is still growing. Response (and representativeness) is decreasing because of lack of commitment with the subject, lack of time of respondents, increasing ‘research’ pressure. Audience research has to be based on large representative random samples to establish reliable reach and frequency figures on large numbers of media titles for relevant target groups. The ongoing increase in quantity produces a (slow) decrease in survey quality. This is a serious threat and asks for a new approach to data collection issues.

2.3 Methodology of EMS

EMS, officially the European Media & Marketing Survey, has proven to be the currency for TV viewership and Print readership among the European elite for the last 15 years. Since the start in 1995 EMS has expanded to measure the top 20% of the population in 22 countries in Western and Central Europe. Since 2006, EMS has expanded to a further eight markets in the Middle East and since 2009 five markets in Africa. The details of the Middle East and Africa surveys are outside the scope of this paper.

The leading principle has always been single source measurement, using data fusion technique (ascription) to solve missing data issues with the profiling questionnaire. The benefit of single source information is the strength of the relationships between audience information, target groups and profiling information. Single source information is however heavily dependent on the cooperation of the respondents and the willingness to provide all this information in an interview or follow-up questionnaire.

The building blocks of EMS in a nutshell are:

- Half year waves (n~9,000), reporting on whole year data to smooth trends and provide large sample sizes for drill down analysis
- Screening questionnaires establish the EMS universe on a top 20% income threshold and the selection of EMS eligible’s for further questioning
- The main questionnaire including some general background variables on business position and flying behaviour
- Media questionnaire for Print, TV and Websites consists on some media imperatives (usage) questions and blocks of media titles (see table 2). Some 175 international titles are asked and additionally for each of the 20 countries some 25 national titles, resulting in 675 titles to be delivered each release.
- For each title questions are asked on awareness, reach past 12 month’s (30 days for websites), recency & frequency of reading, viewing or visiting. Additionally for TV channels yesterday day part viewing is asked to create a virtual diary for media planning
- Marketing questionnaire with target group information, online
- Re-contacts for ad-hoc surveys

Media Type	International	National
TV Channels	20	100
Daily Print	3	165
Weekly’s & Sundays	14	130
Fortnightly	2	5
Monthly Print	8	100
In-Flights	28	
Print-related sites	35	
TV-related sites	60	
Other websites	5	
Total titles	~175	>500

Table 2. titles in EMS

EMS is built on a solid base of proven CATI methodology: calling respondents via RDD sampling, using the screening as an establishment surveyⁱⁱⁱ. By asking some basic characteristics we can derive information on the composition of the EMS population, and at the same time and with the same criteria select willing respondents to complete the media questionnaire. The drawback of this approach is that costs are rising with the decrease in response rates. For EMS screening costs are even rising with a factor of five, because only 20% of the respondents are selected as EMS eligible's.

With new media types (websites and mobile platforms) and the increase in media titles the length of the questionnaire has increased in one decade from 25 minutes to 35 minutes but the response has dropped by almost 50%.

It is clear that only adding more questions, without any further change in the methodology, would mean the end of EMS.

2.4 EMS 2.0

Although probability sampling is still the leading paradigm in research, questions can be raised on the selectiveness of fixed telephone sampling frames, as used in RDD, especially in combination with the increased non-response. In 2007 EMS started to do experiments with online interviewing as a replacement for CATI sampling but realized that a full transition to an online EMS was not feasible. The crucial bottleneck is the fact that online panels are not indisputable, and therefore not acceptable as the basis for a currency, and the absence of a unified definition of our universe, the European Elite, in all European markets. In 2008 the project EMS 2.0 was formulated, with the following goals:

- Future proof for another decade
- A good balance between quality and costs, and high response is of course is the best guarantee to optimize both
- No change in universe definition: top 20% households, top 13% income earners
- Representative net samples and a good representation of several hard-to-reach groups of respondents (business people, mobile only)
- Every wave stable and reliable R&F results

Although the project title was EMS 2.0, the feeling was that principles of Web 2.0 (no one-way (asking) questions but co-creation and building social networks) was a bridge too far. At the moment it is still necessary to remain within the paradigm of old-school survey methodology of probability sampling and good response rates, but adding components that would enhance the quality of the survey, such as mixed mode data collection and data fusion. There was confidence in finding a workable solution:

- The affluent population of EMS is highly digital
- Respondents, even from the affluent society, are more willing to cooperate in compact surveys of about 15 minutes than on a comprehensive survey of more than 30 minutes
- Interviewers are more committed to do convince respondents to cooperate when they know the burden will be acceptable
- Advanced weighting, fusion and calibration techniques can complement the paradigm of probability sampling (and in the future possibly replace).
- And finally: there was no real alternative. To continue with response rates dropping, costs rising, and questions on data quality growing was a dead end.

For establishing actual population information on the affluent universe CATI (probability) screening is still a necessary component, which can't be replaced by online data collection alone. Within the screening phase of EMS the necessary description of the EMS eligible population is established, which in turn is used in the weighting process.

3. Methodology of EMS 2.0

The new methodology was based on three pillars (figure 1):

- Reduce the length of the questionnaire by asking only half of all media titles to every respondent with a random block design
- Use data fusion to create complete data sets by filling in the missing media information
- Expand the sample size with 50% more interviews from online data collection in a mixed mode design

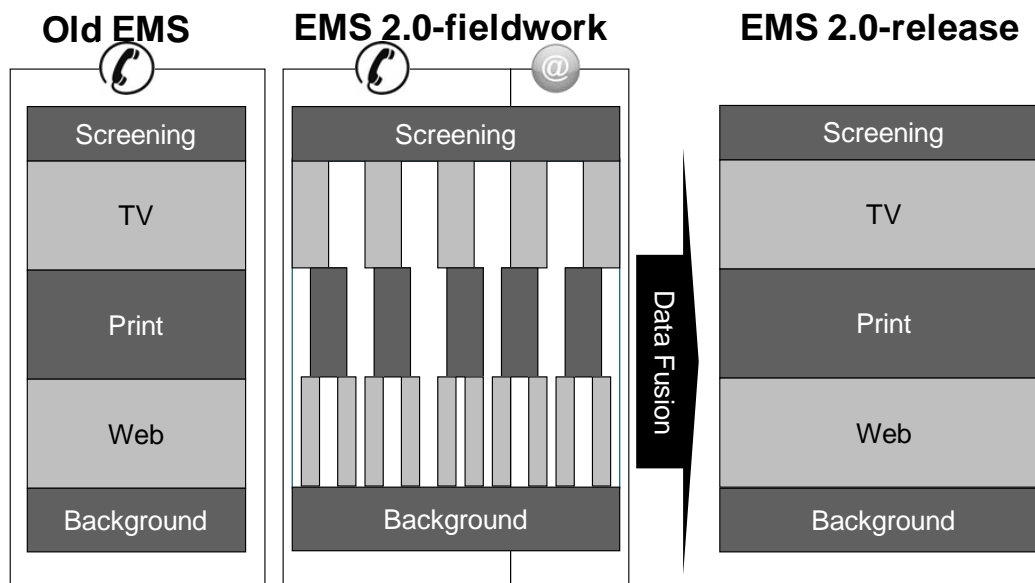


Figure 1. change in methodology of EMS

In the new design the screening interviews still play an important role in the universe definition and the weighting schemes. To improve the quality of the response and the data, the length of the questionnaire has to be limited drastically and as a result sample sizes have to be increased to get comparable reliability levels:

- Sample sizes are large enough to establish the levels of audiences, even for specific sub-samples
- The socio-demographic variables, which play an important role in the data fusion, are asked of all respondents. They help to transfer the information to the missing answers
- In each release the missing information is always a quarter of all available information. Two waves of data are compiled to one database, and only the new wave has to be fused, the previous wave already having been fused in the previous release. So the missing information in the data fusion is only 25% of the total sample.

After testing the new design on the 2008 fieldwork, the new methodology was introduced in the 2009 fieldwork and released to clients from 2010.

3.1 Random block questionnaire design

The core of the new EMS is a random block design. The EMS media questionnaire has a block structure which is rotated randomly. Twelve versions (figure 2), with different combinations of blocks, only 50% of all media titles questions are asked in such a way that data fusion later on can reproduce all levels of and relations between the different media titles and relevant target groups. General questions as screening, travel and socio demographics are asked to everyone.

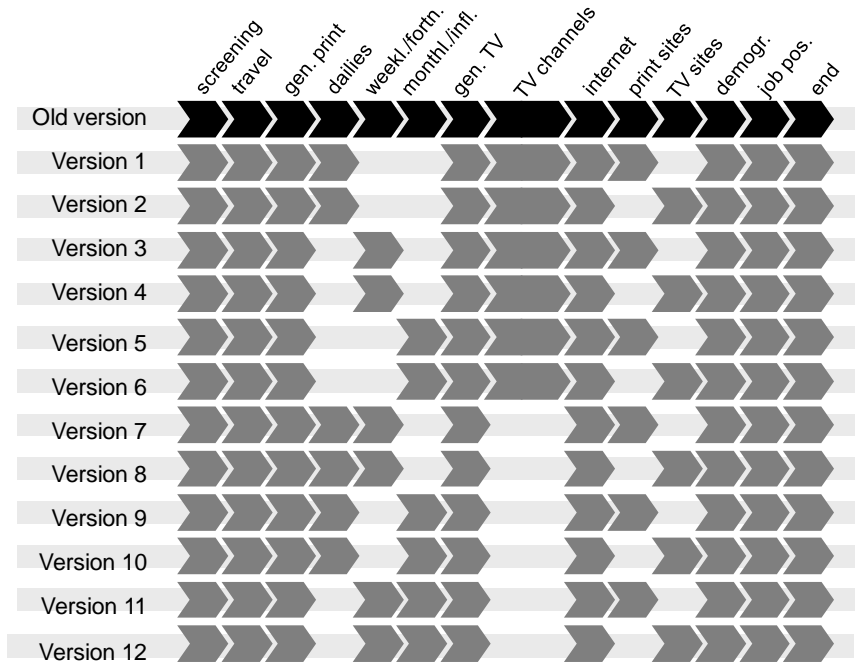


Figure 2. Random block Design

The block with TV-questions was kept together because of the diary block of questions on yesterday day part viewing. Versions 1-6 get all TV questions, while version 7-12 only get the general TV questions. Print is split in three blocks: dailies, weeklies & fortnightlies and monthlies & in-flight-magazines. Websites were split in two blocks: Print related and TV related sites. All combinations of titles are available in this design, to give direct information on relations between titles.

3.2 Chaid based data fusion

Data fusion in audience research is an accepted technique for missing data imputation. The most familiar method is ascription. Ascription is based on a distance function; a complete record of information is captured from the donor and transferred to the recipient. This technique has the advantage that within the record of information copied all relations and values are consistent and real (existing), but the disadvantage is that it relies on the strength of the relation between ‘hook’ (mostly socio-demographic) variables and all the other variables of interest. Sometimes socio-demographic variable X is a good ‘predictor’ for media variable Y, but not for Z. It is impossible to find a set of hooks which are good ‘predictors’ for all media (target-) variables.

Alternative fusion techniques are often model based and called imputation. The advantage is that for each target variable a model is used to analyse the relationship of the target variable with the other variables in detail to reproduce this relationship. There are also some disadvantages (and a whole literature to fix these^{iv}), like the problem of non-existing values, high frequency of one imputed value, assumptions on (linear) relations, and so on. For EMS 2.0 a combination of ascription (hot deck) and model based imputation has been developed: Chaid based data fusion.

For each media title a Chaid tree is built, modelling to the most discriminating splits for the donors. From the end-nodes, when the groups are too small, or no significant split is possible anymore, donors are ascribed to recipients with the same characteristics. This last (hot deck) ascription within the end-nodes of the Chaid tree is done using the weights of respondents to ‘marry’ donors with a large weight to recipients with a large weight.

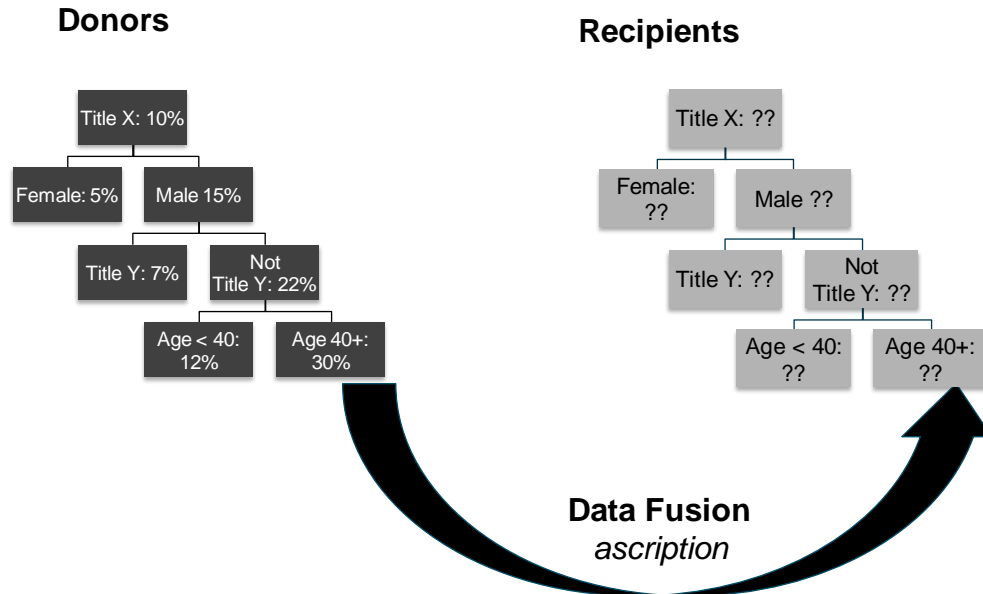


Figure 2. example of a Chaid based data fusion

In figure 2 this is worked out in a fictitious example: Most discriminating factor for title X is gender, for male we need to split between male respondents who read title Y and those who didn't. The last group can be split into age groups. The same tree is built for the recipients and the information from the end node of the donors is given to the recipients

This Chaid based data fusion has several advantages:

- All levels of audiences are respected: end-nodes get - on average - the same level of audience
- Ascription on end-node level copies the same variation in answer patterns: instead of assigning everyone within the end node with exactly the same value, the complete distribution and range of values of the donor are copied
- Chaid takes into account all possible relations (non-linear, linear), with socio-demo's, but also with other titles as well
- Chaid based data fusion corrects for small deviations from random distribution of the question blocks within the random block design. End nodes are constructed (with significantly different values from other end nodes) and copied from donor to recipient, but the size of these end-nodes is subject to weighting. If for example title X is asked of too few female readers, the size of this female group will be adjusted in the weighting, but they will get the right value in the data fusion process.

Without prior ideas on the structure of the data, the Chaid analysis find those differences which can be significant later on, and with data fusion we construct the same structure back. These principles of Chaid based data fusion may look simple, the practice is not:

- More than 30,000 Chaid trees need to be analyzed, because we have almost 700 titles, with several variables in iterative runs
- Different variables per title are subject to data fusion, which are all inter-connected and have to be run sequentially: first total reach (for Print and TV past 12 month, for website past 30 days), then frequency and recency of reading, viewing and visiting. As well as other variables such as awareness of titles, place of viewing, read/viewed on trip were taken into the runs as well
- For each variable several iterative runs were executed: first making trees only on socio-demographic variables to get an initial value, then including titles from the same media type, and finally other media type titles were included in the Chaid trees which could have over 200 end nodes.
- Computing time takes several days rather than hours, which made testing alternative options a very time consuming exercise

3.3 Larger sample size and mixed mode data collection

Larger sample sizes are needed when less questions are asked. 50% additional (online) respondents were added in the fieldwork creating a mixed mode design. For EMS Mixed mode data collection was never the ultimate goal or solution, but a practical one: can we increase sample size for affordable costs and as addition to the traditional Cati fieldwork in terms of relevant target groups. Besides that it also brings extra sources of possible errors: can we distinguish between the effect of the different sample sources (panels) we use and different mode effects?

In the last couple of years the discussion of using online data collection has been dominated by all kind of issues around panel quality. For EMS it was clear from the start that online data collection is a necessary (and cost effective) addition to correct for the selectiveness of CATI sampling but not a replacement of it. Some target groups, such as mobile only households cannot be reached by fixed line sampling alone. The European EMS population is highly digital, with a penetration of over 90% in most countries (figure 3).

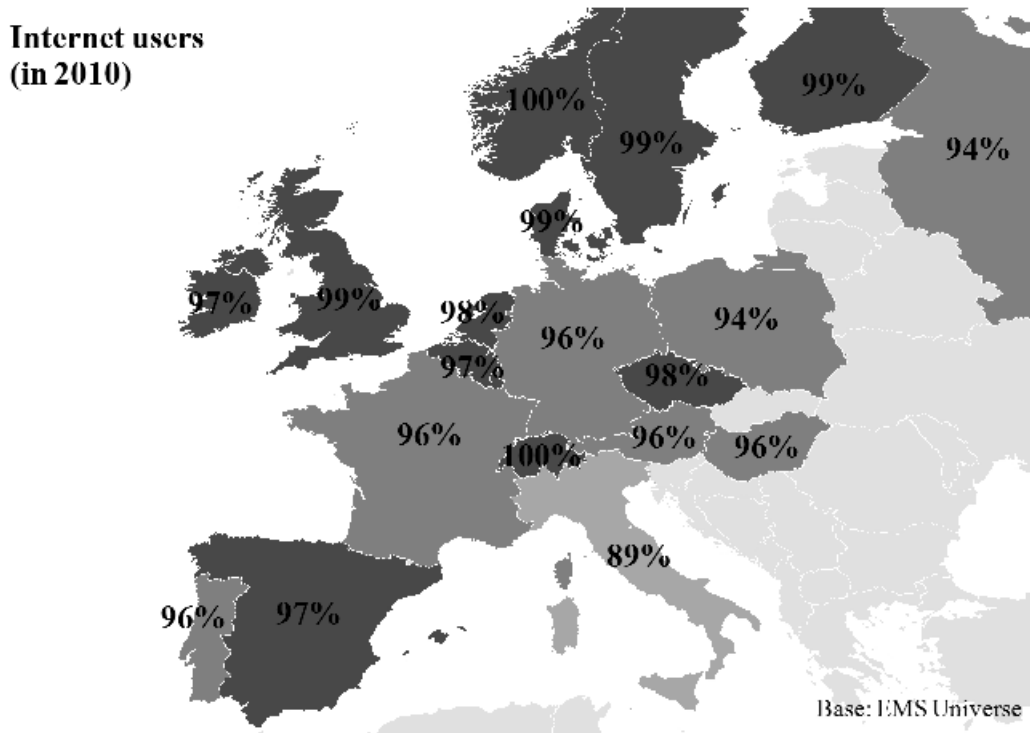


Figure 3. internet penetration within the EMS population 2010

The online fieldwork has been outsourced to external vendors, and is based on quota sampling (quota's on some basic characteristics of the eligible EMS population per country as income, gender and age). Comparing relevant (business) characteristics between the CATI and online fieldwork shows that differences are relatively small and within acceptable wave-to-wave sampling fluctuations. The general picture of the distribution of most variables is the same (see for instance main activity, table 3). Online we have however less self-employed and less 60+ hours working per week (table 4).

Unweighted	CATI	Online
Main activity		
1. Finance	7.9%	8.8%
2. Other services	18.6%	20.3%
3. Professional	6.4%	8.9%
4. Healthcare or medicine	10.6%	8.2%
5. Manufacturing, engineering & construction	21.1%	18.6%
6. Education	8.7%	7.0%
7. Government, excl. education + healthcare	6.2%	7.1%
8. Police or Armed forces	1.2%	3.0%
9. Agriculture or mining	1.0%	0.7%
10. Other	18.4%	17.5%
Total	100.0%	100.0%

Table 3. Main activity CATI-online-unweighted, base fieldwork 2009 wave 1

Unweighted	CATI	Online
Occupation/ position		
1. Self-employed	18.6%	14.6%
2. In paid employment	70.7%	74.9%
3. Temporarily not working	0.8%	1.1%
4. Retired	9.5%	8.8%
5. Not gainfully employed	0.3%	0.7%
Total	100.0%	100.0%
Working hours		
0. None	10.6%	10.5%
1. less than 20 hours	1.8%	2.2%
2. 20-31 hours	6.1%	4.7%
3. 32-39 hours	12.9%	21.1%
4. 40-59 hours	53.3%	54.1%
5. 60+ hours	15.2%	7.4%
Total	100.0%	100.0%

Table 4. Occupation & working hours CATI-online-unweighted, base fieldwork 2009 wave 1

To directly access the EMS eligible’s the online fieldwork can be seen as a good alternative for CATI probability sampling and screening. When it comes to the composition of relevant target group characteristics, sometimes it is even a necessary addition (for instance for mobile only households). However comparing online and CATI fieldwork on so called ‘media imperatives’ shows that the online fieldwork has a different media consumption pattern (table 5): more heavy web users and TV users, slightly more light print users. These differences in media imperatives also have an impact on the audience measures. To correct for this side-effect in the online fieldwork, the weighting process was expanded, to incorporate media imperatives in the weighting.

Unweighted	CATI	Online
Media imperative Web		
1. Light (0-4 hpw)	49.4%	13.6%
2. Medium (4-8 hpw)	25.3%	20.4%
3. Heavy (8+ hpw)	25.3%	66.0%
4.		
Total	100.0%	100.0%
Media imperative TV		
1. Light (0-7 hpw)	30.7%	20.7%
2. Medium (7-14 hpw)	34.9%	25.5%
3. Heavy (14+ hpw)	34.4%	53.8%
Total	100.0%	100.0%
Media imperative Print		
1. Light (0-3 hpw)	28.7%	31.2%
2. Medium (3-6 hpw)	34.1%	33.8%
3. Heavy (6+ hpw)	37.2%	35.0%
Total	100.0%	100.0%

Table 5. Media imperatives CATI-online, base fieldwork 2009 wave 1

3.4 Weighting and smoothing

The survey data are weighted on a national basis and grossed to the universe estimate of the affluent EMS population by country. The weighting is conducted on one complete year database, two waves of fieldwork. The weighting is carried out for each country / market in several steps, from household information based on the random screening sample to the main income earners information for all samples within the EMS population.

All random interviews within EMS eligible households are analysed and transformed in to a database of the top 13% main income earners of all 22 markets, corresponding to the top 20% households with the highest income. The weighted data is the base for establishing the sample distribution for the main socio demographic variables such as gender, region, household size, number of main income earners and age groups.

With the new methodology the distribution for media imperatives (Print, TV and Web usage) is also established on the random fieldwork to incorporate the online sample in the weighting scheme, and corrected for the skews of media behaviour of the online fieldwork. This is the same procedure as is used to incorporate a ‘business boost’, extra sample from calling companies to reach specific business functions.

EMS is released twice a year, on a rolling two wave database. The Data Fusion is done only on the last wave of fieldwork, the weighting on the whole year. This causes a minor smoothing effect, because the data fusion copies the audience levels and data structure from donors from the last (half) wave and the previous (full) wave, which is partly fused itself.

4. Results

In theory the new approach looks fine: randomly left out questions are replaced with information from comparable respondents with the same level of answering. Before changing the methodology however, testing on a complete data set was necessary to make sure that this approach would not harm the credibility and reliability of EMS.

The most important quality indicator is of course the response rate (corresponding to the questionnaire length), which is still the best guarantee for data quality. This was the ultimate goal for EMS 2.0, to achieve a cost effective and high quality instrument for affluent audience research for the next decade.

The final judge on the validity of this approach is the consistency of the trend line.

4.1 Pre-testing data fusion fieldwork 2008

On an existing data set (EMS fieldwork 2008) 50% of the answers on the media questions in the second wave of 2008 were left out and reproduced by Chaid based data fusion as described above. This setup resembles the real situation from 2009, with one great difference: the sample size is not enlarged. So the test is more critical than the real life situation from 2009.

A major advantage of this test above the situation after implementation of the new approach is that now the real answers are known and can be compared to the fused data. We can compare:

- Deviation of the donors from the real value, due to the random block design (and distribution of weights or sample sizes). This deviation is heavily dependent on sample size and penetration of titles.
- Deviation of the donors and the fused value with real value. Data fusion should reproduce the found donor value, but it sometimes also reproduces/corrects for non-random selection of the produced missing values if these are Missing at Random (MAR).

Print

In table 6.1 and 6.2 for the different groups of print titles we see that the average value from the donors sometimes deviate from the real value, but that the data fusion corrects (a little) for this (small) deviation (for example daily print). The data fusion is able to reproduce the donor value correctly and deviations move in the right (real value) direction.

Print Titles		Real	Donors	Fused
Daily	12 months	8.1%	7.8%	7.9%
	AIR	0.4%	0.4%	0.4%
Weekly	12 months	12.4%	12.3%	12.4%
	AIR	2.4%	2.4%	2.4%
FN	12 months	2.6%	2.6%	2.6%
	AIR	0.5%	0.4%	0.4%
Monthly	12 months	4.1%	4.1%	4.2%
	AIR	1.6%	1.6%	1.6%
Inflights	12 months	8.4%	8.4%	8.4%
	AIR	1.9%	1.9%	1.9%

Table 6. Data fusion print titles

TV

For TV we see the same results as for Print. For total reach (viewed last 12 months) the data fusion has a tendency to deviate in the wrong direction: the fused audience levels are higher than the donor information and the real value. This deviation is still within the normal reliability levels and due to the size of the weights in the data. Exact reproduction of audience level is almost impossible, because the random block design is a random process producing random fluctuations, and (large) weights are responsible for some more fluctuations.

TV channels	Real	Donors	Fused
Aware	58.3%	58.3%	58.4%
12 months	32.1%	32.2%	32.6%
30 days	22.1%	22.3%	22.3%
Week	12.9%	13.1%	12.9%
Day	3.5%	3.6%	3.5%

Table 7. Data fusion TV titles

Individual titles on 8 target groups & combinations

As soon as we drill down to smaller target groups and combining reach of different titles more deviations occur: sometimes the random block design produces small deviations from the real value, but they are mainly within the normal reliability levels. As seen by above Data fusion repairs these differences when the produced missing data patterns are MAR.

We tested the number of significant differences for international Print and for 8 different target groups: total, male, UK & Ireland, Nordic, C-EMS, Southern Europe, Top3%, Business Flyers (see table 7).

For TV we see the same results as for Print, as well as on an average group levels as on individual title base.

No. Significant differences		No. Tests	Real - Donors	Donors - Fused	Real - Fused
Print	Recency	297	0%	0.3%	0%
	Frequency	297	0%	0%	0%
TV	Recency	164	0%	1.2%	0%
	Frequency	164	0%	0%	0%
Websites	Recency	208	0%	0%	0%
	Frequency	208	0%	0%	0%

Table 8. Data fusion individual titles

Combinations of titles

For TV 8,000 tests with combinations of different titles could be executed with almost no significant differences. For Web & Print we see some more significant differences in about 5% of the tests, due to the fact that titles are measured in different random blocks.

Conclusion of the test

When these tests were conducted, before the introduction of the new design the results were promising enough to continue. On a total level for the main titles, the picture was very positive, but it was clear that the smaller the target group and the more titles involved deviations occur. It was also clear that after increasing the sample size with 50%, as was decided already, the results would further improve.

4.2 Data quality: response and questionnaire length

Until 2008 questionnaire length increased by 30% and response rates have decreased by more than 40% compared to 2005. Partly due to more titles in the questionnaire, but also because every year it is getting harder to get respondents to cooperate in surveys, so more time is needed to convince people to continue.

From 2009, when the new methodology was introduced, the length of the questionnaire reduced to a level below 2005 (figure 4). The response rate also recovered, but still not to the same level as 2005 because the overall response trend is still going down.

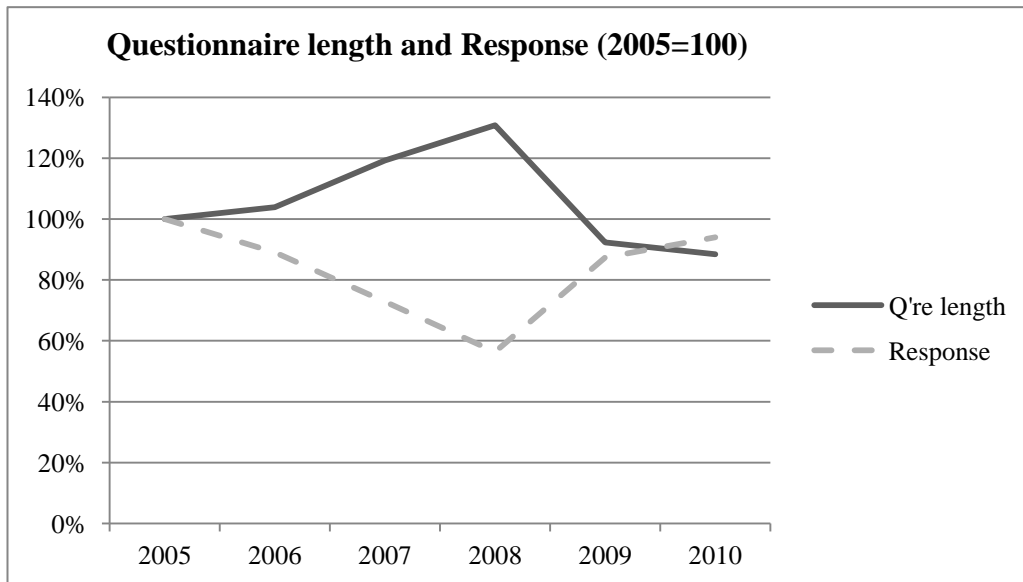


Figure 4. development of questionnaire length and response

With a shorter questionnaire and higher response some important benefits are accomplished: less costs, less respondent fatigue, less pressure on memory/recall capabilities, more and better answers and more interviewer commitment.

4.3 Consistent trend line

The proof of the pudding is in the eating and looking at trend lines for many individual titles it is hard to discover a break indicating the new methodology. After careful consideration with the subscribers EMS decided to promote the new methodology as the start of a new trend, and not as the continuation of the old.

Avoid false claims

The new methodology was the first major change in 15 years and with the new methodology a new stable trend is established. We had to be sure that no claims on trend development from users of the data were subject to the new methodology rather than the real trend.

Other changes

Introducing a new methodology made more survey changes possible and every change may impact on the results. The opportunity to make essential changes in a ongoing media survey are scarce, so this was the moment to make such changes.

Online mode effect

Mixed mode data collection raises issues on sampling effects (different kind of respondents) and mode effects (such as other answering patterns due to the difference between title and logo recognition). It is hard to differentiate between the effects, the online fieldwork is both a replacement and complementary to the CATI fieldwork. Differences between CATI and online can be due to different respondents or different answering patterns. We could correct for part of the online effect with an adopted weighting scheme. However we have no guarantee that the remaining differences are real, mode effects or trends. We have therefore made a fresh start, and with the online component kept constant in the upcoming waves of EMS, we do not have to solve any mode effect, but accept it as the new trend.

4.4 Implementation

EMS is a syndicated survey and involvement of the clients is crucial. Comprehensive tests have been conducted to convince the clients that these steps are necessary and do not impact the results negatively. The results prove that the new approach is the

right one: response increased and is of higher quality, online data collection is smoothly fitted in and data fusion is an elegant and efficient method for the creation of robust and reliable databases.

The new approach has been executed with the support of all the EMS clients (over 40 international media owners and media agencies). They have fully accepted the migration and the EMS results are vital for their (commercial) business performance, which proves the successful outcome of the transition.

5. Conclusion

In this age of information we want to know more from as many consumers as possible, but respondents are no longer willing or able to fill in long and boring questionnaires. EMS has made a start by working with fragmented data by randomly leaving out half of the media questions. Additional online fieldwork, necessary to increase sample sizes, has been integrated smoothly into the process.

Testing the new approach on an existing (and complete) data set proved that data fusion based on Chaid trees is an efficient and elegant way to reconstruct the full data set.

Improving the measurement of a currency without any change is impossible. EMS needed to improve the methodology and the subscribers accepted the change. We are all confident that the new methodology with shorter questionnaires provides higher quality data. Larger sample sizes and data fusion are needed to reconstruct (and reproduce) the data in a reliable and efficient way. So we ask less and get more and better data.

References

ⁱ The effect of Panel Recruitment and Management on Research Results, Nopvo 2006, Esomar Seminar on Panel Research, Barcelona 2006

ⁱⁱ Van Massa naar Mens, Doran van Veenendaal & John Faasse, Swocc Publication 2010

ⁱⁱⁱ EMS essential guide, and EMS website www.ems.synovate.nl

^{iv} Statistical Analysis with Missing Data /Roderick J. Little & Donald B. Rubin, Wiley 2nd edition, Wiley Series in Probability and Statistics