

## **PAMCo SLA Technical Appendix 2017 Report**

### **Introduction**

Item 5.6 of PAMCo Service Level Agreement reporting requires Ipsos to provide an annual technical appendix report one month after delivery of annual datasets (data for January-December 2017 were delivered to the data bureaux on 15 March).

### **Overview of Methodology**

PAMCo – Audience Measurement for Publishers is the new audience measurement currency for published media. It produces de-duplicated brand reach allowing users to carry out reach & frequency planning and also plan and trade audiences across all the platforms on which published media content is delivered – phone, tablet, desktop and print.

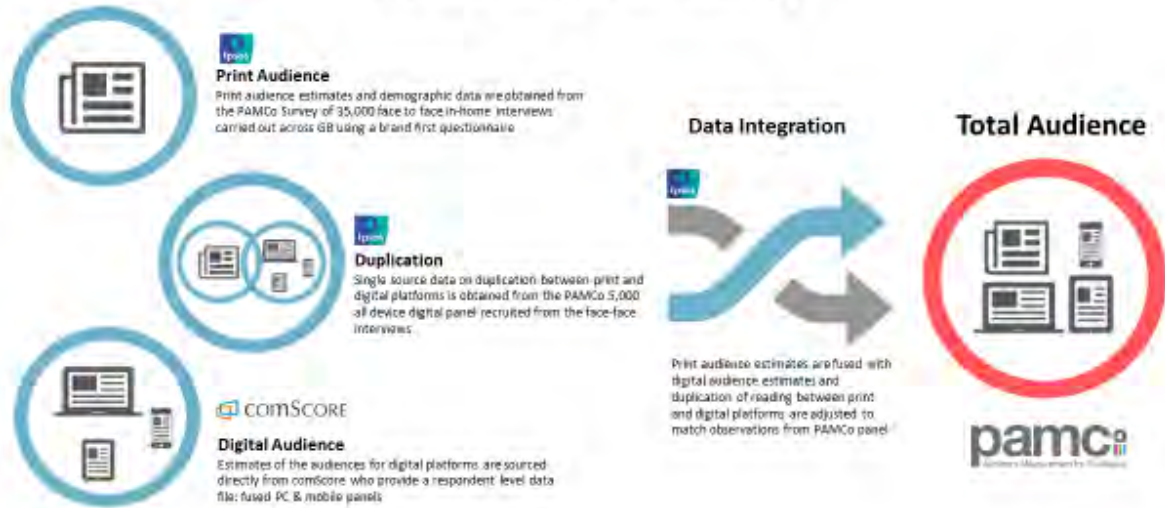
The PAMCo currency is platform neutral and provides the following:

- De-duplicated reach and frequency for all platforms
- An increased number of brands reported across all platforms
- Single source data to understand duplication between print and digital
- Improved estimates of net brand reach and duplication
- Reporting newsbrand sections across print AND digital
- New engagement data

PAMCo's world-leading methodology integrates data from three sources to provide a complete view of publisher audiences.

1. A high-quality face-to-face survey of 35,000 participants per year interviewed in home to collect print readership and demographic data.
2. A digital panel of around 5,000 participants (recruited from the survey participants at the time of interview) to have a specially designed Tracker App installed on their devices i.e. desktop, laptop, phone, tablet. The principal purpose of this panel is to provide single source data on duplication between reading in print and reading on digital platforms. It is not the intention to report digital platform estimates directly from the panel.
3. A new specially developed data file from comScore, UKOM's chosen digital audience supplier, which makes it possible to integrate fully 'phone and tablet audiences as well as PC/laptop audiences (see p.42 for a description of how the file is created). A highly sophisticated data integration methodology has been developed to fuse comScore's estimates with the readership/enumeration survey. Duplication of reading between the print and digital platforms is adjusted to match observations from the PAMCo panel. Duplication of reading between the digital platforms is matched to comScore.

## World Leading Methodology



## Report Contents

This report includes the following sections:

### Survey

<b>Page</b>	<b>Section</b>
3	The Sample
10	Fieldwork and Response Rates
12	The Interview
27	Weighting
29	Modelling and Analysis

### Data Integration

<b>Page</b>	<b>Section</b>
32	The Panel
34	Fusion report
42	Description of comScore's Multi-Platform Respondent Level File

## THE SAMPLE

### The Population Sampled

The sample is designed to be representative of the adult population aged 15+ of Great Britain. The survey population is estimated to have been approximately 52,748,000 according to JICPOP's mid 2017 projections based on 2011 census data.

The target sample size is 35,000 per annum and the achieved sample for 2017 was 34,910 (of which 34,896 were included in the reporting sample).

### Overview of the PAMCo Sample

PAMCo has a multi-stage disproportional pre-selected sample design.

The three stages of sampling are as follows:

- 1) Selection of sampling points (2568 points for 2017)
- 2) Selection of addresses
- 3) Selection of individuals

The sampling frame for the selection of sampling points comprises all Output Areas (OAs) in Great Britain. In 2017, two OAs were selected for a sampling point.

The sampling frame for the selection of addresses is the Postal Address File (PAF).

Interviewers were issued with the pre-selected addresses to contact. At each household contacted (see below for the household definition), they enumerated eligible individuals aged 15+. The final stage of the sampling was then to select which of these eligible individuals should be interviewed. One individual was selected in most households, and up to two individuals in larger households. The rules as to how this was done are described later in this document.

Some disproportional sampling was used at the stage of selecting sampling points in order to over-sample certain ACORN areas particularly likely to contain individuals with AB social grades.

### Household definition

For PAMCo, a household is defined as:

'One person or a group of people who have the accommodation as their only or main residence'  
AND

'For a group of people a household should have some form of common living arrangement – that is they share some of the living accommodation'

Shared living accommodation may include:

- a living room
- a dining room
- a living or dining room which is part of the kitchen
- it would still count if the living room has to be used as a bedroom

However, a group of people is not counted as a single household solely on the basis of a shared bathroom / toilet.

Single addresses may be sub-divided into separate dwellings which are split into separate households for interviewing purposes. In Scotland, tenement buildings are common and when interviewers encounter these they follow the procedure for multi households.

Defining if an address is a multi-household can be tricky and the interviewer is careful to consider:

- if the dwelling is capable of being self-contained, **and**
- if the dwelling is being used by separate households

A property can be considered to be self-contained if:

- It has a separate entrance from the street and has its own kitchen and bathroom facilities
- or
- It has a joint entrance from the street but has a separate individual entrance within a property and also has its own kitchen facilities inside this doorway
- or
- It has a joint entrance from the street but has a separate individual entrance within a property. The kitchen facilities may be shared but there should be no other common living or dining areas (note that a kitchen which has any seating for eating meals would count as a common living/dining area and so would not count as a self-contained dwelling)

## Selection of Sampling Points

The first stage of sampling is to select the sample points (or locations) where interviews will be carried out. Before this is done, certain ACORN areas are upweighted (and others downweighted). This is done so as to oversample areas where individuals with AB social grades are particularly likely to live, and therefore increase the number of interviews taken with AB participants.

The steps taken to select the sampling points were as follows:

- a) Sampling weights were applied to the adult populations of OAs by 7 categories of ACORN Types, as follows:

Category	ACORN Types	Weights Applied
1	1, 2, 3, 9, 14	1.90
2	4, 5, 7, 8	1.65
3	6, 10, 15, 16, 17, 18	1.35
4	11, 12, 13, 19, 20, 21, 22, 24, 26, 29, 61, 00	1.10
5	23, 25, 27, 28, 30, 31, 32, 33, 34, 36, 37, 50	1.00
6	35, 38, 39, 40, 41, 42, 43, 49, 53, 54, 55, 60, 62	0.73
7	44, 45, 46, 47, 48, 51, 52, 56, 57, 58, 59	0.63

- b) Great Britain was divided into 60 area groupings, devised for PAMCo. Ultimately these areas were individually weighted and used to piece together the Government Office Region and non-overlap ISBA TV regions.

A full list of the 60 areas can be found overleaf. For the purposes of sampling, the area groupings were further sub-divided into district groupings.

- c) The weighted adult populations of each of the 60 area groupings were calculated. The target number of sampling points for each area grouping was then set. The total number of points at this stage was 2,568 per annum.
- d) Within each area grouping, OAs were ranked in descending order by district grouping, then within district by ACORN category, and finally within ACORN category by NS-SEC. Prior to 2005, a system of AB scores was used instead of NS-SEC as the final means of ranking.
- e) The required number of OAs within each area grouping as specified at b) was then selected with probability proportional to weighted adult populations, using random start/fixed sampling interval procedures.
- f) Any selected OA falling into a 'no-go' area (as separately defined) was replaced with one in the sample District and same ACORN Type with population size closest to the originally selected OA.
- g) Modified procedures were applied in Scotland in areas 49, 50, 51, 52, 56, 57, 58 and 59. In these areas the target number of points was increased, relative to other areas, to the numbers shown in the table below. This is to allow for larger analysis samples for smaller regional newspapers in Scotland.
- h) Sampling Areas

The area groupings used and the target number of Sampling Points selected were as follows:

## England and Wales

Area Groupings	Government Office Region	ISBA Area	No. of Points Selected
1	South West	South West	72
2	South West	Southern	36
3	South East	Southern	72
4	South East	Southern	36
5	South East	Southern	48
6	South East	Southern	84
7	South East	London	72
8	East of England	London	84
9	Greater London	London	48
10	Greater London	London	96
11	Greater London	London	72
12	Greater London	London	72
13	East of England	East of England	60
14	East Midlands	East of England	24
15	South East	London	36
16	East of England	East of England	36
17	East of England	East of England	60

18	South East	Midlands	48
19	South West	Midlands	24
20	West Midlands	Midlands	72
21	West Midlands	Midlands	48
22	Greater London	London	48
23	West Midlands	Midlands	24
24	West Midlands	Midlands	24
25	West Midlands	Midlands	48
26	East Midlands	Midlands	36
27	East Midlands	Midlands	36
28	East Midlands	Midlands	36
29	Wales	Wales and the West	12
30	Wales	Wales and the West	36
31	Wales	Wales and the West	48
32	North West	Border	24
33	South West	Wales and the West	24
34	South West	Wales and the West	60
35	Yorkshire and the Humber	Yorkshire	84
36	Yorkshire and the Humber	Yorkshire	48
37	Yorkshire and the Humber	Yorkshire	36
38	East Midlands	Yorkshire	48
39	Wales	Lancashire	12
40	North West	Lancashire	60
41	North West	Lancashire	60
42	North West	Lancashire	24
43	North West	Lancashire	48
44	North West	Lancashire	48
45	Yorkshire and the Humber	North East	24
46	North East	North East	48
47	North East	North East	24
48	North East	North East	24

## Scotland

Area Groupings	Area Names	No. of Points Selected
49	Aberdeen	24
50	Dundee	12
51	Edinburgh	36
52	Strathclyde, urban	96
53	Strathclyde, rural	24
54	Lothian	12
55	Central	12
56	Fife	24
57	Other Tayside	24
58	Other Grampian	24
59	Highland*	24
60	Scottish Borders	12

\* Includes one sample point per year for each of the Western Isles, Orkney and Shetland.

## **Allocation to Month and Start Day**

The selected sampling points were systematically allocated to month and then start day, to ensure a balanced daily interviewing schedule. Interviewers were given ten days to complete a standard assignment, starting on the specified day.

## **Selection of Addresses**

Within each selected pair of OAs, 26 or 27 standard and 12 replacement addresses were selected, apart from London where 29 standard addresses were selected. There was a further variation in the number of addresses issued as follows:

In London, the standard number of addresses issued was reduced if there were indications that a high proportion of multi-household/tenement addresses might be found.

With effect from October 2017, addresses were issued to the interviewer via the Electronic Contact Sheet (ECS) software on their tablet computer. Prior to October, hard copy Contact Questionnaires (one for each address) were issued to interviewers.

Interviewers were allowed to use replacement addresses only when they identified that an original address was a:

- Business address
- Addresses not located, empty or demolished
- Non-main residence

Having recorded the original address as one of the above codes, the ECS selected a replacement address to be used.

A minimum of five calls was prescribed for each potentially productive address, before attempts to interview could be abandoned. There was no restriction on the total number of calls. Interviewers were encouraged to make as many calls as possible, varying them by time of day.



## Selection of Individuals

The selection of individuals was done within the ECS using the following steps:

- a) All selected addresses within each selected OA were allocated the letters P, Q, R and S consecutively.
- b) For all issued addresses the ECS specified the one or two persons to be interviewed, as follows:

### Address Letter

No. of adults at address	P	Q	R	S
1	First	First	First	First
2	First	Second	First	Second
3	Second	First, Third	Second, Third	First
4	First, Fourth	Second, Third	First, Fourth	Second, Third
5	Second, Fourth	First, Fifth	First, Third	Second, Fifth
6+	First, Youngest	Third, Fifth	Second, Fifth	Fourth, Youngest

The ECS instructed interviewers to list all persons aged 15+ in a selected household, in descending order of age and the ECS then confirmed the person(s) selected for interview, depending on the total number of adults at the address and the letter allocated to that address.

At multi-household addresses, tenements (in Scotland) and residential institutions, there were special selection procedures, details of which are available on application to Ipsos MORI.

## FIELDWORK AND RESPONSE RATES

The results on the following page are for the January-December 2017 fieldwork period and are based on interviews carried out between 16 December 2016 and 3 January 2018. Fieldwork results are shown for original and replacement addresses separately (first and second columns) and for all addresses in total (final column).

**Known to be away temporarily** means that the interviewer found that the person to be interviewed would be away from home for the whole of the interviewing period in that area.

**Out, 5 or more calls** means that the person to be interviewed was known to be 'in residence' but was out on each of the five or more occasions that the interviewer called.

**Refused** includes all refusals both by and on behalf of the person to be interviewed.

**Interview not possible** means that the person to be interviewed was incapable of giving a satisfactory interview because of sickness, infirmity, deafness or other such incapacity, or inability to speak English.

**Other Reasons** include any other reasons not included in the above descriptions.

The PAMCo response rate is a complete response rate because it reflects both of the following:

- The proportion of the sample that is successfully contacted in the first place (77.9% in 2017)
- The proportion of identified participants who complete an interview (59.6% in 2017).

Multiplied together, these two elements provide the overall response rate, which in 2017 was 46.4% (target 46.5%).

- Participant incentives

Participants in the London region and a defined area within the South East of England (areas which present the most challenges when attempting to achieve an interview) received a £20 incentive (Love2Shop giftcard) for a completed CAPI interview, or could opt for Ipsos MORI to donate this sum to one of three designated charities on their behalf.

## PAMCo January – December 2017

	Original addresses		Replacement addresses		All addresses	
Total addresses issued	66737	100%	5592	100%	72329	100%
<u>Less</u> business addresses	610		80		690	
<u>Less</u> address not located/empty or demolished	4310		404		4714	
<u>Less</u> non-main residence	411		68		479	
	5331	8.0%	552	9.9%	5883	8.1%
<u>Plus</u> additional multi-household generated	962		91		1053	
<u>Less</u> empty household	89		9		98	
A) Total potential contact households	62277	100%	5124	100%	67401	100%
<u>Less</u> No reply received at any call	8402		568		8970	
<u>Less</u> Contact refused	5574		382		5956	
	13976	22.4%	950	18.5%	14926	22.1%
Contact interview achieved	48301	77.6%	4174	81.5%	52475	77.9%
B) Selected persons	53890	100%	4643	100%	58533	100%
<u>Less</u> No reply received at any call	3928		266		4194	
<u>Less</u> Known to be away temporarily	1287		84		1371	
<u>Less</u> Out, 5 or more calls made	728		82		810	
<u>Less</u> Known to be busy temporarily	973		75		1048	
<u>Less</u> Refused	14127		890		15017	
<u>Less</u> Interview not possible	730		39		769	
<u>Less</u> Other reasons for no interview	395		22		417	
Number of interviews obtained with selected persons	31722	58.9%	3188	68.6%	34910	59.6%
Overall response rate (A x B)		45.7%		55.9%		46.4%

## THE INTERVIEW

This section provides details of the interview for PAMCo that has been deployed from January 2017. Key points regarding the interview are summarised below.

The interview is carried out on CAPI using a single screen tablet, which means that the interviewer and the participant share the same screen.

PAMCo is a brand-first interview. Readership questions start by presenting brands in groups of six and asking whether any of the six have been seen in the past 12 months. Then the individual brands that have been seen in the past 12 months are established before confirming whether the brand is read on any print or digital platform. This is then followed by more detailed questions about separate reading on print and digital platforms.

PAMCo uses full colour masthead prompts for the readership questions, including non-date specific front covers for the magazines and newspaper supplements.

PAMCo asks specific questions about weekday and Saturday newspapers, showing separate weekday and Saturday paper front covers.

The PAMCo interview adapts the frequency scales used to the different types of publication:

Magazines	Almost Always (3 or 4 copies out of 4)
	Quite Often (2 copies out of 4)
	Occasionally (1 copy out of 4)
	Less often
Weekday newspapers	Almost Always (4 or 5 copies a week)
	Quite Often (2 or 3 copies a week)
	Occasionally (1 copy a week)
	Less often
Saturday and Sunday newspapers	Almost Always (3 or 4 copies a month)
	Quite Often (2 copies a month)
	Occasionally (1 copy a month)
	Less often

### Average Issue Readership

The principle of measuring print 'readership' is to attempt to establish at each interview whether or not the person interviewed has looked at any copy of the publication in question during a period back from the day of interview equal to the interval at which the publication appears. (Reading on the actual day of interview is not included). Thus, for each daily newspaper claimed to have been read in print in the past 12 months, the interview attempts to establish whether or not the person interviewed looked at a copy of it 'yesterday', the day before the interview (or 'yesterday or Saturday' in interviews conducted on Mondays). For each Saturday/Sunday newspaper or weekly magazine read in the past 12 months, the interview finds out whether or not the participant has looked at a copy of it in the past seven days, and for each monthly magazine,

whether or not he or she has looked at a copy of it during the past four weeks. Other publication frequencies include fortnightly, bi-monthly and quarterly magazines.

## Sections of the Interview

The PAMCo interview is divided into five main sections which are presented to the participant as follows:

1. What you read (the main part) – including topics of interest, internet usage, readership
2. What devices you use (if any) – a section covering **internet devices** and **activities and frequency of internet usage**.
3. General information about your household, including classification
4. What other media you consume (TV, Radio, Cinema)
5. Joining our panel (if eligible) – panel recruitment and recontact section

The following outlines the content of each section.

### SECTION 1 (Introduction and Readership)

The topic interest question appears at the start of the interview and presents the participant with 30 topics. The screen has the following prompt:

“I would like to ask you about the topics that you read or look at in a newspaper or magazine whether in print, on a computer, on a tablet or on a mobile phone.”

This section also asks the participant about future plans and internet usage.

#### The readership section

Throughout the readership section, the interviewer shares their tablet with the participant.

The readership section, which uses a brand-first methodology, establishes which publications the participant has read in the past 12 months, whether in print or on screen, followed by more detailed questions on those that have been read in the past 12 months. Readership could be in the form of printed newspapers or magazines, or newspaper or magazine websites/ apps that have been looked at on a PC, tablet or mobile device.

Publisher brands are initially shown in groups of six and the participant determines whether he/she has read or looked at or any one or more from the group in the past 12 months. For the groups that are selected, the participant will be shown those groups again later and asked which brand(s) from the group he/she has read in the past 12 months.

The interview uses media list rotations to prompt brands as fairly as possible across the media list. The rotations for displaying publisher brands in groups of six are as follows:

Forward: Daily Newsbrands, Regional Daily Newsbrands, Magazines, Sunday Newsbrands

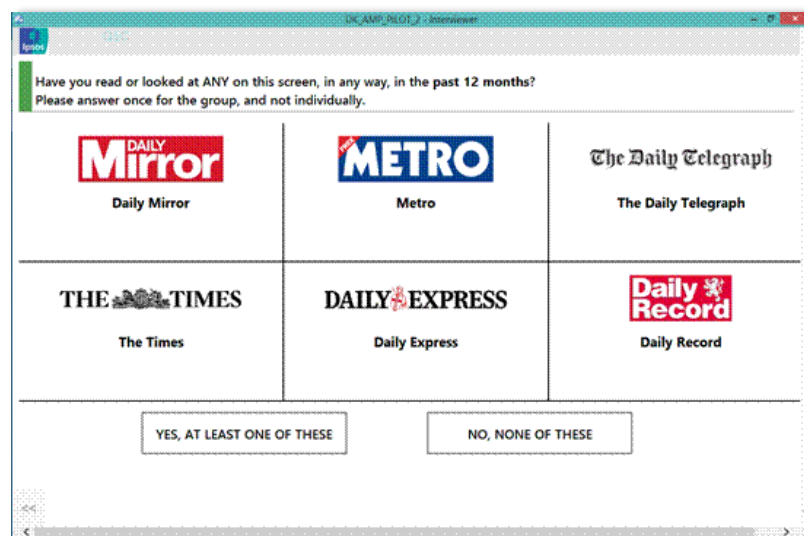
Reverse: Sunday Newsbrands, Magazines (in reverse order), Daily Newsbrands, Regional Daily Newsbrands

The order in which brands are displayed on-screen in groups is also rotated randomly, again to be fair to all brands. For a full summary of the media list rotations see page 21.

The interview then goes on to ask the participant whether the magazines or newspapers were read as a printed copy or on-screen via a website or app.

Example screenshots from the interview are below, to show how the questions appear on screen.

- Publisher brands (groups)



- Yes or No for each brand in groups selected



- Print or On screen for each brand selected



The screenshot shows a survey window titled 'UK\_PAMCO\_TEMPLATE\_V2 - Interviewer'. The survey question is: 'Have you read or looked at The Guardian in print or on screen in the past 12 months? Please answer yes or no for in print and on screen.' Below the question is the 'theguardian' logo and the text 'The Guardian'. There are two sections: 'IN PRINT' with 'YES' and 'NO' buttons, and 'ON SCREEN' with 'YES' and 'NO' buttons. A green 'NEXT >>' button is at the bottom right.

The readership section also establishes whether the participant has looked at websites or apps where there is no equivalent printed magazine, or the printed magazine does not have the same name as the website or app.

Given that the objective of PAMCo is to provide estimates for readership in print and on screen, the frequency of looking at the publication on-screen is asked first (if the brand has been claimed as looked at on screen in the past 12 months), before moving on to ask about when the publication was last read in print and then how often it is read in print. There is evidence from NRS (the predecessor to PAMCo) that some claims for print readership are in fact for on-screen readership, so by asking about on-screen readership before print readership we are aiming to minimise incorrect claims.

If both magazines and newspapers have been claimed as read in the past 12 months, magazines are asked about first before asking about newspapers.

## Magazines

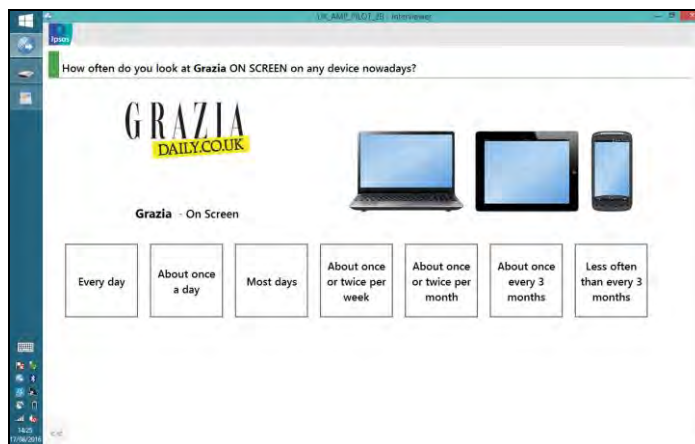
The recency ('when last') and frequency ('how often') questions are prompted on a brand-by-brand basis in the example sequence shown below for the weekly magazine Grazia.

Grazia - On Screen (how often)

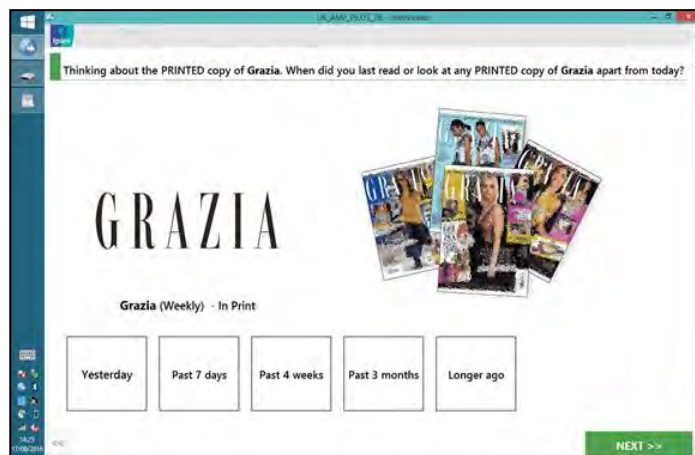
Grazia - In Print (when last)

Grazia - In Print (how often)

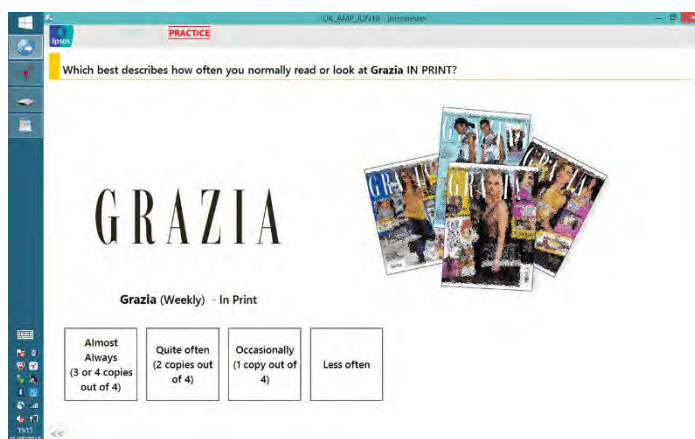
On Screen (how often):



In Print (when last):



In Print (how often):



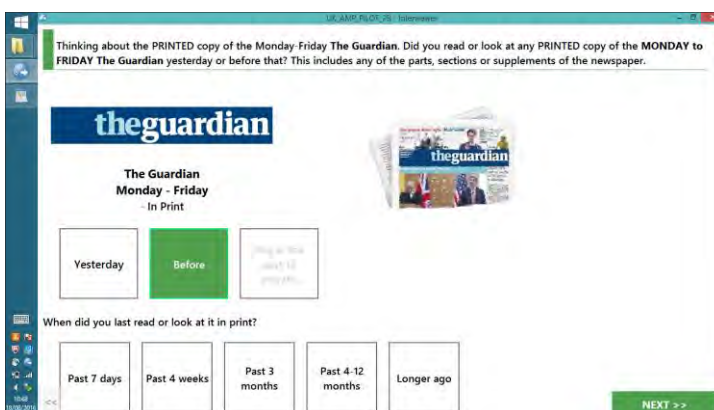


## Newspapers

For newspapers the different issues or editions of a newspaper are prompted consecutively, with the supplement and in-paper sections right at the end. An example of the sequence for a participant who claims to have read both the weekday and Sunday editions of The Times newspaper in the past 12 months, and The Times on screen, is shown below.

- The Times – On screen (how often)
- The Times (Monday-Friday) – In Print (when last)
- The Times (Monday-Friday) – In print (how often)
- The Times (Saturday) – In Print (when last)
- The Times (Saturday) – In Print (how often)
- The Sunday Times – In Print (when last)
- The Sunday Times – In Print (how often)
- The Times (Saturday supplements) – when last
- The Times (Saturday in-paper sections) – when last
- The Sunday Times (Sunday supplements) – when last
- The Sunday Times (Sunday in-paper sections) – when last

The Monday-Friday ‘when last’ question is asked in two stages as shown below for the Guardian.



Throughout the readership questions the words ON SCREEN and IN PRINT are emphasised to help the participant to think about the different platforms. The wording for the different newspaper issues (Monday-Friday/Saturday/Sunday) are emphasised for the same reason.

As well as asking about readership of printed newspaper supplements i.e. stand-alone magazines or other separate sections that come with a newspaper, in-paper sections for some newspapers are also prompted i.e. sections that are part of the main newspaper itself. An example screenshot is show here:



## Local Weekly Newspapers

If the sample point is in a postcode where a local weekly newspaper is measured, questions are asked in the interview about print readership in the past 12 months, 'when last read' and how 'often read'. These data are not published by PAMCo but are supplied to JICREG as an input for their modelling procedures.

- Engagement questions

There are 3 different engagement questions in the interview which are:

- Source of Copy
- Time Spent Reading
- Attitude Statements

## Source of Copy

A source of copy question is asked following completion of recency and frequency questions for all titles and is asked for print average issue readers only.

Participants are shown a list of precoded options as follows:

- I obtained it myself
- Someone else in my household obtained it
- Someone from another household obtained it
- Saw work or office copy
- Saw copy elsewhere

The participant is then asked follow-up questions which vary depending on the publication's distribution method (paid/free, paid only and free only).

The questions are as follows:

Did you or your household buy it, or did you get it for free?

- 1 Paid
- 2 Free
- 3 Don't Know

Which of these best describes how the last copy was obtained?

- 1 Bought at newsagent or shop
- 2 Delivered by newsagent
- 3 Delivered with online shopping
- 4 Delivered directly from the publisher
- 5 Through subscription by post
- 6 Subscription vouchers from publishers
- 7 I subscribe to it
- 8 Picked up on bus, train or tube
- 9 Picked up on a plane or at airport
- 10 Picked up in a shop
- 11 Other
- 12 Don't Know

Which of these best describes how the copy was obtained?

- 1 On a bus, train or tube
- 2 On street or at a station
- 3 Picked up in a shop
- 4 At a recreation outlet (gym, club etc)
- 5 At a student campus
- 6 Other
- 7 Don't Know

### **Time Spent Reading**

A 'time spent reading' question is asked after the source of copy question for print average issue readers only. Participants are shown a list of precoded options and asked "*Roughly how long do you usually spend in total reading or looking at an issue of (NAME OF TITLE) by the time you've finished with it, including all the times you look at it and all the parts and sections?*"

The participant's answer is coded into one of the following codes:

- Less than 5 minutes
- About 5 minutes
- About 10 minutes
- About 20 minutes
- About 30 minutes
- About 45 minutes
- About 1 hour
- About 2 hours
- About 3 hours or longer

For daily newspapers, time spent reading is recorded separately for the weekday and Saturday issues, with an instruction to include readership of all the sections of the newspaper.

### **Attitude Statements**

The following statements are asked if either print AIR and/or digital frequency 'every day' or 'most days' has previously been claimed in the PAMCo interview:

- Reading it is time well spent
- I feel a close connection with it
- It gives me something I can't get elsewhere
- I trust what I read in it

The participant is asked to confirm one of the following responses for each statement:

- |                |                   |
|----------------|-------------------|
| Agree strongly | Disagree a little |
| Agree a little | Disagree strongly |

## Media list rotations

The following table summarises the order of questions for newsbrands and magazine brands and the rotations applied in order to treat all brands as fairly as possible.

Question	Forward Rotation	Reverse Rotation
Q5C: Brands in Groups of Six (any in past 12 months, whether print and/or on-screen). Random rotation of position of each brand on the screen.	Daily Newsbrands Regional Daily Newsbrands Magazines Sunday Newsbrands	Sunday Newsbrands Magazines (in reverse order) Daily Newsbrands Regional Daily Newsbrands
Q5E: RPY of Individual brands seen in past 12 months, from groups selected above	The order selected at Q5C	The order selected at Q5C
Q5G: Platform (print/on-screen) for brands selected for RPY at Q5E	Titles within publication types in the order selected at Q5E. Exception is magazine brand with a shared website, which appear consecutively (with the first of the brands appearing in the order selected at Q5E).  Daily Newsbrands Regional Daily Newsbrands Sunday Newsbrands Magazines	Titles within publication types in the order selected at Q5E. Exception is magazine brand with a shared website, which appear consecutively (with the first of the brands appearing in the order selected at Q5E).  Magazines Sunday Newsbrands Daily Newsbrands Regional Daily Newsbrands
Individual brand questions (recency and frequency), magazines followed by newsbrands for both rotations.	Magazines (in order selected at Q5G):  Q7A: on-screen frequency Q8A: Print recency Q8B: Print frequency  Newsbrands (in order selected at Q5G). When both a daily and a Sunday edition of the same brand are selected at Q5G (Platforms), the brand will appear in the order of whichever of its editions were prompted first at Q5G, while still maintaining the question order overleaf	Magazines (in order selected at Q5G):  Q7A: on-screen frequency Q8A: Print recency Q8B: Print frequency  Newsbrands (in order selected at Q5G). When both a daily and a Sunday edition of the same brand are selected at Q5G (Platforms), the brand will appear in the order of whichever of its editions were prompted first at Q5G, while still maintaining the question order overleaf

	<p>Q9A: On-screen frequency            Q10C: Print recency (Mon-Fri)            Q14: Print frequency (Mon-Fri)            Q10E: Print recency (Sat)            Q13: Print frequency (Sat)            Q15A: Print recency (Sun)            Q18A: Print frequency (Sun)            Q11A: Print recency (Saturday supplements)            Q12A: Print recency (Saturday in-paper sections)            Q16A: Print recency (Sunday supplements)            Q17A: Print recency (Sunday in-paper sections)</p>	<p>Q9A: On-screen frequency            Q10C: Print recency (Mon-Fri)            Q14: Print frequency (Mon-Fri)            Q10E: Print recency (Sat)            Q13: Print frequency (Sat)            Q15A: Print recency (Sun)            Q18A: Print frequency (Sun)            Q11A: Print recency (Saturday supplements)            Q12A: Print recency (Saturday in-paper sections)            Q16A: Print recency (Sunday supplements)            Q17A: Print recency (Sunday in-paper sections)</p>
Q6A: Digital only brands	Magazines Newsbrands	Newsbrands Magazines
Q6B/Q7B: RPY and frequency	In order selected at Q6A	In order selected at Q6A
Q19: Local weekly newsbrands	Print title prompted by postcode on a sample point basis.	Print title prompted by postcode on a sample point basis.
Engagement Questions – each question asked for all brands before moving to next question.	<p>Q20A: Source of Copy (Print AIR)            Q20F: Time Spent Reading (Print AIR)            Q21A: Attitude Statements (Print AIR or equivalent digital frequency)</p>	<p>Q20A: Source of Copy (Print AIR)            Q20F: Time Spent Reading (Print AIR)            Q21A: Attitude Statements (Print AIR or equivalent digital frequency)</p>
Q20A: Source of Copy (Print AIR). Titles within publication types in the order selected at recency and frequency questions. If AIR of both a daily and Sunday edition of a newsbrand have been selected, they appear consecutively.	Daily Newsbrands Sunday Newsbrands Magazines	Magazines Daily Newsbrands Sunday Newsbrands

<p>Q20F: Time Spent Reading (Print AIR). Titles within publication types in the order prompted at Q20A Source of Copy.</p>	<p>Daily newsbrands (Mon-Fri/Sat)            Daily newsbrands (Mon-Fri only)            Daily newsbrands (Sat only)            Sunday newsbrands            Magazines</p>	<p>Magazines            Daily newsbrands (Mon-Fri/Sat)            Daily newsbrands (Mon-Fri only)            Daily newsbrands (Sat only)            Sunday newsbrands</p>
<p>Q21A: Attitude Statements (Print AIR or equivalent digital frequency). Titles within publication types in the order prompted at Q5G Platforms. If AIR of both a daily and Sunday edition of a newsbrand have been selected, they appear consecutively.</p>	<p>Newsbrands: Print/Digital Mon-Fri/Sat            Newsbrands: Print/Digital Sunday            Magazines: Print/Digital Digital only</p>	<p>Magazines: Print/Digital            Newsbrands: Print/Digital Mon-Fri/Sat            Newsbrands: Print/Digital Sunday Digital only</p>

## SECTION 2 (Internet devices activities and frequency of usage)

This section covers the different types of devices that can connect to the internet. The participant is asked if he/she has any of the devices listed below in the household and how many he/she uses personally. In a single person household, the questions for 'number in household' are suppressed.

- Computer
- Smartphone
- Mobile
- E-book reader
- Tablet
- Smart watch
- Video games console
- Smart TV
- Blu-Ray disc player
- Streaming device
- Set top box or DVR

These questions are presented on two grids, the first of which is shown below.



The screenshot shows a survey question interface with the following text: "Now thinking about devices that can connect to the Internet. Please tell me how many of each of these you have in your household and then tell me how many you personally use. Please only count those that are WORKING AND IN USE." The interface is divided into six sections, each with a device category and two input fields: "HOUSEHOLD" and "PERSONALLY USE".

Device Category	Household Count	Personally Use Count
Computer (exclude work computers)	2	1
Smartphone	0	0
Mobile	0	0
E-Book Reader	0	0
Tablet	0	0
Smart Watch	0	0

A "NEXT >>" button is located at the bottom right of the interface.

Work computers are excluded at this question, and are defined as computers that are owned by the employer for purposes of work. If the participant is self-employed and/or owns their own computer, the computer should be classified as a household computer.

As well as ownership of devices, the participant is asked how many of a particular device he/she owns, plus time spent using these devices.

### SECTION 3 (Classification information)

The classification section starts with questions about shopping and then moves on to demographic questions about the participant, including gender and ethnicity. Then there are questions about the number of children and adults in the household.

The gender question in the script is as follows:

"Which of the following describes how you think of yourself? [Use showcard]"

Male

Female

In another way

Prefer not to say

Next more detailed information about the occupation of the Chief Income Earner and participant are collected. Not only are occupational details required for social grading purposes, but the industry and job are also coded in their own right for SIC and SOC (see below).

The classification section concludes with questions on income and education.



## The Chief Income Earner (CIE)

The Chief Income Earner (CIE) can be of either gender and is determined by asking the following question:

*“Which member of your household would you say is the CIE, that is the person with the largest income, whether from employment, pensions, state benefits, investments or any other sources?”*

The Chief Income Earner must be related to the participant. The Social Grade of unrelated people living in the same household is based on their own individual occupations. Couples living as married/co-habiting are treated as ‘married’ and therefore ‘related’ for classification purposes. Widows receiving maintenance, with no higher grade occupations, are graded on their late partner’s occupation. If two or more people in the household have equal incomes, then social grade is coded for the older person.

## SIC and SOC codes

SIC (Standard Industrial Classification) codes are government definitions used to classify a company's nature of business or the type of economic activity in which it is engaged. It is a very extensive list, for example manufacturing alone may be one of 258 different codes depending on exactly what is manufactured.

SOC (Standard Occupation Classification) codes are another government classification. Like the SIC codes it is a very long list, for example a teacher may be one of 44 different codes and an engineer one of more than 200.

The interviewer probes for the following information in order to code social grade, SIC and SOC accurately:

- Exactly what type of firm the CIE/participant works for – is it retail, manufacturing etc. What does the firm sell, make?
- The exact job title
- What he or she does – what does he or she actually do in his/her day-to-day role.
- The number of employees he or she is responsible for and how many work there
- The qualifications that he or she holds for his/her position – if a formal qualification this should be the level (HND, degree, O level etc.)

## Social Grade

Once all of the relevant details to enable social grading have been recorded by the interviewer, a summary of the information is shown on screen for the interviewer to code social grade. All social grade coding is checked once fieldwork has concluded by experienced coders at Ipsos MORI.

If the CIE is unemployed i.e. he/she doesn’t have a paid job but is actively seeking, or is available for, employment at the time of interview, then the following rules are used for social grading purposes:

If, at the time of interview, the unemployed person is dependent on state benefit only, the social grade is based on his/her previous occupation if he/she has been unemployed for less than six months.

If the unemployed person who is dependent on state benefit only has been unemployed for over six months at the time of interview, his/her social grade will be E.

If the unemployed person has income exceeding state benefit level, the social grade will be based on his/her previous occupation regardless of the length of unemployment.

### **Retired participant or Chief Income Earner**

If the CIE is retired, additional questions are asked to help code social grade accurately.

There is a question on whether he/she has a private or company pension or not. Someone receiving a private pension, however small, is coded as having a private or company pension.

When talking about his/her occupation, if retired then he/she is asked about his/her main or highest occupation. If it has already been established that he/she has a private pension then this is the occupation which they are classified on.

If the pension is from a late partner's occupation then this is the occupation that is used for grading.

The interviewer confirms if the CIE has benefits only or also has another form of income. This other form of income may include, but is not restricted to, a company pension, a widow's pension, and monies from savings or investments. If the CIE is retired and has no other income apart from a State pension then the social grade is E.

If the CIE is retired and has any form of income over and above state benefits, the social grade is based on his/her former occupation.

### **SECTION 4 (Other Media)**

This section covers how often the participant goes to the cinema, listens to the radio and watches the television.

### **SECTION 5 (Panel Recruitment)**

At the end of the interview, eligible participants (those who have claimed primary usage of one or more eligible devices earlier in the interview) are asked to join the PAMCo panel, with a target of 5,000 participants recruited to the panel from the 35,000 interviewed. A full panellist successfully installs specifically designed software, which passively measures digital consumption of published media, onto all their eligible devices. Eligible devices are PCs, smartphones (Apple and Android) and tablets (Apple and Android). The purpose of the panel is to provide single source data on duplication between reading in print and on digital platforms.

A full panellist received £10 if they had 1 eligible device installed, £15 for 2 eligible devices and £20 for 3 or more devices. The incentive took the form of a Love2Shop giftcard or donation to one of three designated charities.

## WEIGHTING

For 2017, the following weights were applied to each quarter's data:

- a) Pre-weights to correct for disproportional sample selection by ACORN Type Category.
- b) Pre-weights for household size.
- c) Pre-weights for 60 area groupings
- d) Social grade smoothing to produce exponentially smoothed estimates of target proportions for social grade within sex. The factors applied to the current quarter's figures/previous estimates are 0.25/0.75.
- e) Rim-weighting by sex within age group, sex within area grouping and sex within social grade. At this weighting stage the 15-24 age group was split into 15-17 and 18-24 age groups.

The sample was grossed to JICPOPS mid 2017 estimates (2011 census based), a total of 52,748 (000s).

The rim weighting efficiency, highest/lowest weights and range of weights by quarter are shown below.

<b>Summary of Weighting Results: 2017</b>				
	<b>Q1 2017</b>	<b>Q2 2017</b>	<b>Q3 2017</b>	<b>Q4 2017</b>
Rim weighting efficiency	88.9%	89.8%	87.7%	89.8%
Highest weight	5.4482	5.4125	5.0796	4.6951
Lowest weight	0.0941	0.1217	0.0798	0.1105



## MODELLING AND ANALYSIS

For 2017, 34,910 interviews were achieved, of which 34,896 were included in the reporting sample for PAMCo 1 2018. 14 interviews were not included in the reporting sample further to the following coding at the gender question:

8 participants coded 'Prefer not to say'  
6 participants coded 'In another way'

With effect from the January 2018 fieldwork period, participants with the above codes will be included in the reporting sample.

### **Modelling Procedure for Daily Newspapers**

A modelling procedure was applied to data for daily newspapers.

The PAMCo data show results for three AIR measures as follows:

Monday-Friday average issue readership (weekday AIR)  
Saturday issue readership (Saturday AIR)  
Monday-Saturday average readership (6 day AIR)

#### *Monday-Friday average issue readership (weekday AIR)*

The PAMCo interview asks about Monday-Friday readership of daily newspapers and Monday-Friday AIR is captured directly from the interview data and not modelled.

#### *Saturday issue readership (Saturday AIR)*

These are obtained from a specific question on the readership of Saturday issues of national daily newspapers.

#### *Monday-Saturday average readership (6 day AIR) and read past year (6 day RPY)*

AIR targets are calculated from 1/6 of the Saturday ('Yesterday' or 'Past 7 days') and 5/6 of the Monday-Friday ('Yesterday') AIR.

6 day Monday-Saturday Read Past Year (RPY) readership needs to be calculated for reporting purposes and this is done by creating nets of the weekday and Saturday editions for both recency and frequency data.

When netting recency, the more recent of the Monday-Friday and Saturday intervals is chosen where a participant is a reader of both editions. For example, if a participant claimed last 7 days recency for the Monday-Friday edition and last 4 weeks for Saturday edition, then the participant is assigned the 'last 7 days' recency for the Monday-Saturday newspaper.

There is no obvious choice for how the frequency net should be calculated. However, considering that the Monday - Friday edition is 5 out of 6 issues of Monday-Saturday, it is logical that if a person is a reader of the Monday-Friday edition, the Monday-Friday is used, without regard to Saturday frequency. If a person is not a reader of the Monday-Friday edition, a frequency of "Occasionally" or "Less Often" is given ("Less Often" for every first and second occurrence and "Occasionally" for every third occurrence).

The AIR modelling procedure described above was applied to the AIR in Government Office Region Scotland separately from the AIR for the rest of Great Britain for 5 national daily newspapers. These are newspapers which have a different masthead for their Scottish editions: Daily Express, Daily Mail, Daily Star, The Sun and The Times.

## Probabilities

PAMCo readership data are collected and processed in a binary format. Each participant is either a reader (1) or not a reader (0) of a publication. In order for PAMCo readership data to be used by the data bureaux in reach and frequency planning tools it is preferable that PAMCo binary data are converted to probabilities. Probabilities give each participant claiming to have read a publication in the past year (RPY) a score of between 0 and 1 of being an Average Issue Reader (AIR) of the publication. Probabilities are derived using a CHAID segmentation technique which is based on demographic and other variables available from the interview data.

## Probability of Reading for Newspaper Supplements and In-Paper Sections Print Data

In the PAMCo interview, there is no frequency data collected for newspaper supplements and in-paper sections, only recency data. To allow the calculation of cumulative readerships and of schedules, an alternative method of calculating probabilities has been derived.

All participants that claimed to read the supplement or in-paper section in the past 12 months are segmented according to their propensity to be average Issue readers of the supplement or section. Segmentation is conducted using CHAID. Once segmentations are complete, the terminal nodes of the segmentation are split according to frequency of reading. As there are no frequency data for the supplements and sections, the frequency of reading the parent newspaper is used as a proxy, and an AIR probability is calculated for each frequency node.

Each participant has, for each supplement or section, an individual 5-digit probability code ranging from 0 to 1 e.g. 0.4828. These probabilities do not correspond to the standard frequency codes of other publications.

Summary codes of the probabilities are provided and are defined as follows:

High	0.6667-1.0000
Medium	0.3334-0.6666
Low	0.0001-0.3333

## 24 Month Base Procedure

The 24 month base procedure allows smaller titles to have PAMCo published estimates. The procedure is applied to titles falling below the annual threshold of 140 unweighted average issue readers (AIR) required for publication of estimates, but achieving an unweighted AIR of 180 or more over a 24 month period.

The procedure combines data from 24 continuous months of data, including the current quarter, to calculate the combined penetration of each of the recent reading and frequency of reading claims in the total population. Readership targets based on the most recent 24 months are then overlaid on the most recent 12 months of data, with any adjustments necessary made to the most recent quarter. This may have the effect of adjusting read past year (RPY) as well as AIR.

A summary of the procedure for adjustment is as follows:

- 1) On an annual basis, identify and specify publications that qualify for this procedure.
- 2) Combine audiences for the most recent quarter and the preceding 7 quarters (21 months) of data.
- 3) Calculate new AIR and RPY targets from the combined population penetrations for frequency and recent reading.
- 4) Adjust weighted data for the current quarter, according to the targets. Randomly select participants to add from the 'non-reader' population, or to remove from the reading population. Selection is based on a target demographic profile, created from the combined quarters' data.

Titles published on a 24 month base have the symbol T in the bureaux systems and in certain PAMCo data tables.

### **Publications with Frequency Changes**

The published readership estimates for titles that change their publication frequency are calculated in the following way:

The average issue readership comprising qualifying recency codes from the new frequency are added to the average issue readership comprising qualifying recency codes from the old frequency.

Publications that have changed their publication frequencies are given the publication symbol F in certain PAMCo data tables to denote that the estimates are calculated using data collected from more than one publication frequency.

### **Adjustment for Circulation Loss**

In addition to the demographic weighting, results are adjusted for the effects of circulation loss where there has been interruption of normal publication. The adjustment procedure involves changing the codes of some participants from 'non-reader' to 'reader' for a relevant title.

Estimates of circulation loss are based on data reported to the NMA or PPA by each publisher together with ABC information concerning National Newspapers. Non-members of the NMA and PPA report directly to Ipsos MORI.

## THE PANEL

Those who have used the internet in the past month and personally use a home computer, smartphone or tablet are invited to join the panel. They are requested to install the PAMCo app on all the devices they use to access the internet. The app only counts the number of visits to publishers' websites and apps and the PAMCo website itself over a period of 4 weeks, with participants requested to keep it installed for 8 weeks to maximise the chances of collecting 4 weeks of "good" data.

A total of 34,896 PAMCo interviews were in the 2017 analysis sample with 78.9% of participants (27,558) eligible for digital panel recruitment.

- 37.3% of eligible participants agreed to join the PAMCo Panel in 2017.
- 26.7% of eligible participants installed the app on at least one of their devices, and 17.3% installed on all devices.
- 27,558 eligible participants had in total 61,055 devices (an average of 2.2 devices per participant)
- Of the full panellists there were an average of 1.8 devices per person.

4,879 full panellists from the 2017 fieldwork period were deemed active and were used for duplication in the fusion. A further 722 from Quarter 4 2016 NRS fieldwork were also used in the fusion for the PAMCo 1 data release.

### **The Panel's role in calibrating duplications**

The PAMCo Digital Panel provides single-source data from PAMCo participants on the overlap between their claimed print reading and their actual visits to publisher websites, passively measured across all of their devices. A passive measure of digital behaviour is required, as claimed digital behaviour is very different from actual behaviour. The panel does not itself generate audience estimates but determines how much more (or less) likely a print reader is to read the same brand in digital than a non-print reader. This information is used to generate a print/digital duplication target which is applied to the fused PAMCo /comScore database.

The measure of duplication is observed at the monthly reach level and it is derived by combining the print readership claim for the last 4 weeks before the main interview and any observed digital reading in the 4 weeks that the participant is tracked on the panel in the period following the interview.

The panel is first weighted in order to correct for skews towards ownership of fewer rather than more devices, towards print readers rather than non-readers of print, and towards women.



Weighting targets are applied for:

- age within gender
- number of newspapers and magazines read in past 4 weeks
- for each title - past 4 weeks print readership by device.

The duplications observed in the weighted panel undergo further modelling in order to become duplication targets, initially at the level of monthly reach across all digital platforms. This is necessary because the digital readership levels in the panel do not match the digital audiences in the comScore currency, and have to be adapted to take account of the size of the comScore digital estimates relative to the size of the PAMCo print estimates.

The modelling used also addresses the issue of titles with low penetrations or tagging issues, resulting in low or zero samples in the Panel which would otherwise prevent a duplication target from being calculated. To overcome this problem, a Bayesian Smoothing approach is used, whereby the duplication target of title X is permitted to partially (or completely, in the case of no sample for a title) rely on how duplication targets for other newsbrands and magazines are distributed. This model also draws on claimed duplication data from the PAMCo survey. Although this information is not reliable per se due to the biases in the recall of digital behaviour, it is highly correlated with a title's true duplication target, and greatly improves the ability to predict duplication targets for smaller brands.

Once the final smoothed duplication targets have been derived for digital overall, a more detailed model is applied to calibrate duplications by platform and platform combination.

## DATA INTEGRATION: FUSION

The following describes the fusion for the first PAMCo data release, PAMCo 1 2018, for the data period January-December 2017. The data inputs are as follows:

- **PAMCo Digital Panel:** October 2016 – December 2017. n=5,601
- **PAMCo Print Survey:** January – December 2017. n=34,896 (28,457 Internet users in past 4 weeks)
- **comScore MMX MP Data:** November 2017.
  - Desktop production sample: n~84,000
  - Smartphone production sample: n~5,600
  - Tablet production sample: n~1,200

### Summary of Fusion Process

The PAMCo fusion is a Multi-step data integration using 3 different data inputs. The first step is a fusion between the PAMCo print and comScore digital audience datasets. Duplication targets are calculated using the Digital Panel, a subset of the total PAMCo sample, and the last step is the calibration to adjust the cross-platform duplication in the fused dataset.

The fusion is a row-wise constrained fusion, in which the print dataset is entirely preserved and the digital dataset is closely matched. Age and Gender are critical cells for the fusion. Both comScore and PAMCo universes are defined as internet users in the last four weeks.

Variables used for matching are:

- Claimed website usage
- Device usage
- *Predicted* Social Grade and Region
- Topic interest

The following process is followed in running the fusion:

1. comScore data preparation - Convert to wide format. Select required websites. Transform demographics to required format.
2. Universe alignment - Align comScore universe with PAMCo universe. Create platform user definitions on comScore and PAMCo
3. Ascription of comScore missing audiences - Aligning comScore file to comScore currency, accounting for “missing” audiences
4. Correction of comScore sectional hierarchy – making sure sections are properly nested within parent title
5. Correction of comScore Mobile/PC overlap – making sure title platforms totals are according to targets
6. Correction of comScore daily data – making sure daily data is within appropriate limits and conforms to targets
7. 15-17 mobile audiences fusion – Creating mobile usage for this age group (not included in comScore data)

8. Group comScore websites potentially covered by the same claimed variables in PAMCo. Group claimed variables in PAMCo potentially relevant for same website and create relevant nets in both files.
9. Align comScore and PAMCo scales – In order to do matching on similar scales (comScore is no. of visits/days, while PAMCo is claimed frequency).
10. Based on various scales, create scores for fusion to potentially work with – binary flags/frequency scale/ number of visits/ functional forms/factors
11. Imputation of Social Grade and Region variable to comScore, based on website visitation and demo profile files
12. Running the Fusion
13. Creating duplication targets
14. Calibrating duplications according to targets

## Quality control

The fusion can be evaluated according to the following criteria:

- Universe estimates within critical cells must be aligned perfectly, subject to universe definition differences between the two databases (Universe Concordance).
- For PAMCo, all results should be preserved exactly at all levels.
- In total and for age and gender (critical cells), the results from ComScore should be preserved (subject to universe differences).
- In matching duplication data to targets the amount of digital data changed should be as small as possible.

### 1. Universe Concordance (Refer to File "Population Comparison AMP Dec'17)

Table 1.1 lists the comScore internet using population by gender and age for November 2017 compared to the PAMCo published internet user population for Jan-Dec 2017.

comScore's population is UK while the PAMCo universe is GB only; the comScore population is therefore downweighted to GB-only targets. Aside from this difference, although comScore uses PAMCo data to weight their universes, there remains a mismatch between PAMCo and comScore audiences due to a number of factors.

- PAMCo populations are higher for Mobile, with a very close match for PC and Total populations.
- Where some universes are lower in the comScore file, (as some people do not use any of the reported websites), additional comScore population is created with no usage, in order to match universes between the two datasets.
- comScore does not provide mobile audiences for 15-17s so an additional fusion algorithm has been developed, the results of which can be found in the reporting tables.

**Table 1.1 – Universe Comparison PAMCo Vs. comScore**

Gender	Age	comScore Internet users (Population)	comScore Internet users (MP File)	PAMCo Internet users	Universe Differences MP vs PAMCo File
Total	Total	43,600	42,529	43,701	3%
Female	15-17	769	993	993	0%
Female	18-24	2,733	2,632	2,652	1%
Female	25-34	4,302	4,144	4,179	1%
Female	35-44	3,842	3,700	3,871	5%
Female	45-54	4,000	3,859	4,080	6%
Female	55+	5,896	5,680	6,054	7%
Male	15-17	807	1,043	1,043	0%
Male	18-24	2,927	2,821	2,804	-1%
Male	25-34	4,266	4,110	4,240	3%
Male	35-44	3,815	3,674	3,768	3%
Male	45-54	3,993	3,848	3,981	3%
Male	55+	6,248	6,025	6,035	0%

## 2. Matching PAMCo Print Currency

Detailed checks of the fusion outputs verify that the Print readership data is completely preserved in the fusion process.

### 3. Matching comScore digital Currency (Refer to Tables “currency Vs. Fused AMP Dec’17, comScore Nov’17” and “Digital Comparison AMP Dec ’17, comScore Nov’17”)

Ipsos receives comScore daily and monthly targets that match the published currency (excluding Northern Ireland) for all aged 18+ for mobile and all aged 15+ for pc/desktop audiences, and the fusion aims to produce digital audiences for each entity that are within +/-3% of the comScore reach, page views and Minutes, at the total audience level and by platform.

Note that 6 large non-publisher entities fell outside of the range +/- 5% on the pc platform in the Q4 2017 fusion and have been suppressed from the dataset on that platform only. Suppressed entities are shown in Table 3.1 below.

**Table 3.1 – Web Entities Suppressed due to PC Platform differences**

comScore entity	Fused Monthly Reach / Currency
Facebook	-8.08%
Facebook And Messenger	-7.88%
FACEBOOK.COM	-7.82%
YOUTUBE.COM	-7.79%
GOOGLE.CO.UK	-7.40%
Complex Video	-5.74%

On average across all entities, the fused total is very close to the comScore target audience for both monthly and daily reach. In Table 3.2, the average difference between fused and comScore monthly reach for all 532 PAMCo reported entities is shown by category.

When looking across categories, on average newsbrands and magazines both fuse to within 0.2% of the comScore monthly reach

Table 3.3 show the entities with the largest difference between comScore and fused monthly reach by category.

**Table 3.2 – Average difference between fused and comScore monthly reach by category**

Category	PAMCo Entities	Difference (fused-comScore) %
Newsbrand	28	-0.2%
Magazine	70	0.2%
Digital only magazine	20	0.4%
Digital only newsbrand	5	0.3%
Digital section	68	-0.5%
Non-publisher	341	0.2%
Total	532	0.2%

**Table 3.3 – Entities with largest difference between fused and target audience by category**

Category	Largest positive difference			Largest negative difference		
	Name	Audience ('000)	% change	Name	Audience ('000)	% change
Newsbrand	PRESSANDJOURNAL.CO.UK	366	0.7%	Daily Record (Mobile App)	30	-2.3%
Magazine	TODAYSGOLFER.CO.UK	83	3.2%	COUNTRYFILE.COM	200	-3.6%
Digital only magazine	KNOWYOURMOBILE.COM	96	2.8%	BIKERADAR.COM	456	-0.4%
Digital only newsbrand	HOMESANDPROPERTY.CO.UK	524	0.4%	INDEPENDENT.CO.UK	19,445	0.2%
Digital section	DailyRecord.co.uk - Homepage	127	2.3%	The Sun Video	106	-2.6%
Non-publisher	PUMA.COM	28	4.5%	VOLVOCARS.COM	160	-4.8%

Table 3.4 summarises the % of published entities that fall within the ranges +/-1.5%, +/-3% and +/-5% in total and by platform.

**Table 3.4 – Matching to comScore estimates**

	% of published entities within +/-1.5%	% of published entities within +/-3%	% of published entities within +/-5%
Monthly reach - TOTAL	90.6%	97.2%	100%
Monthly reach – pc	66.2%	89.7%	96.8%
Monthly reach – phone	100%	100%	98.8%
Monthly reach - tablet	92.5%	98.6%	99.5%
Monthly Duration (Mins)	99.8%	99.8%	99.8%
Monthly Page Views	99.4%	99.6%	99.8%
Daily reach – pc	83.6%	94.8%	97.9%
Daily reach – phone	96.5%	99.6%	100%
Daily reach - tablet	88.0%	97.5%	99.1%

The Excel file “**Digital Comparison AMP Dec ’17, comScore Nov’17**” illustrates that audiences have also been preserved for web entities by platform for the critical variables of age within gender. An example for BBCGOODFOOD.COM is shown in Table 3.5.

**Table 3.5 - Audience comparison by critical variables for example web entity of BBCGOODFOOD.COM**

Website Name	Gender	Age	PC	Phone	Tablet	Total	PC	Phone	Tablet	Total	PC	Phone	Tablet	Total	Fused / Reweighted & Corrected
			Monthly reach MP File	Monthly reach MP File	Monthly reach MP File	Monthly reach MP File	Monthly reach Reweighted & Corrected	Monthly reach Reweighted & Corrected	Monthly reach Reweighted & Corrected	Monthly reach Reweighted & Corrected	Monthly reach Fused	Monthly reach Fused	Monthly reach Fused	Monthly reach Fused	
BBCGOODFOOD.COM	Male	(all)	981	2114	559	3307	948	2170	554	3205	947	2170	553	3205	0.00%
BBCGOODFOOD.COM	Male	15-17	12	0	0	12	11	127	14	143	11	127	14	143	0.00%
BBCGOODFOOD.COM	Male	18-24	90	354	25	445	87	342	24	405	87	342	24	406	0.25%
BBCGOODFOOD.COM	Male	25-34	119	847	37	987	115	818	36	925	115	819	36	925	0.00%
BBCGOODFOOD.COM	Male	35-44	160	258	83	458	159	249	83	407	159	250	83	408	0.25%
BBCGOODFOOD.COM	Male	45-54	225	295	111	509	213	285	105	460	213	285	105	459	-0.22%
BBCGOODFOOD.COM	Male	55+	375	360	303	896	362	348	292	865	362	347	291	863	-0.23%
BBCGOODFOOD.COM	Female	(all)	1304	5246	1097	7282	1259	5494	1069	7357	1260	5493	1071	7359	0.03%
BBCGOODFOOD.COM	Female	15-17	19	0	0	19	19	426	10	444	18	426	10	444	0.00%
BBCGOODFOOD.COM	Female	18-24	161	1215	17	1374	156	1174	16	1303	155	1172	16	1301	-0.15%
BBCGOODFOOD.COM	Female	25-34	192	1691	231	2074	186	1633	223	1995	186	1633	222	1995	0.00%
BBCGOODFOOD.COM	Female	35-44	246	1068	260	1519	246	1032	251	1438	248	1031	251	1438	0.00%
BBCGOODFOOD.COM	Female	45-54	264	785	222	1219	247	758	215	1138	247	758	217	1140	0.18%
BBCGOODFOOD.COM	Female	55+	421	487	367	1076	406	471	354	1040	406	472	356	1041	0.10%

#### 4. Duplication between print and digital (Refer to “MML Integration” and “Calibration Vs. Expected and “AMP Duplication Data Nov’17 including Daily and Weekly)

The file “MML Integration” lists the print and digital vehicles for which duplication is to be controlled in the fusion and works at the level of individual web entities with their respective print titles by platform. Note that where print sections are included within the parent title, these are not separately specified. Digital entities must be tagged in order for the duplication between print and digital to be observed on the PAMCo Panel. The MML file also lists which non-publisher sites are to be matched to the comScore currency in the fusion exercise.

Even though the duplications will later be corrected via calibration, it is important that the fusion is already creating duplications as close as possible to target. Usually, during fusion some of the duplication is lost. We can look at this as a measure of how much better the fusion has reflected duplications compared to random matching.

As shown in Table 4.1, the results of the fusion are in most cases better than random before we further control by the duplications observed between print and digital in the PAMCo panel. Only one digital entity (The Mirror’s Mobile App) has results that are not “better than random”.

**Table 4.1 – Fusion comparison summary for newspapers and magazines**

Sample Bracket (MP File)	Total	Fusion Better Than Random	Fusion Not Better Than Random
Magazines Total	68	68	0
<40	2	2	
40-99	10	10	
100-174	8	8	
175-399	22	22	
400+	26	26	
Newspapers Total	27	26	1
<40	2	2	
40-99	2	2	
100-174	3	2	1
175-399	2	2	
400+	18	18	

A further test of the fusion’s success is how much data needs to be changed via duplication calibration. We expect that no more than 30% of a title’s data is to be changed during the calibration process.

Table 4.2 shows that the proportion of data which had to be changed to bring duplications to target levels was not high for most entities. (See “Calibration Vs. Expected Dec’17). For 94%, we needed to change less than 10% of data. The one entity with greater than 30% changed is The Guardian (Mobile App).

**Table 4.2 - % of digital data changed for fused entities**

% of digital data changed for print titles	Fused Entities	% of total
>30%	1	1%
10-29%	5	5%
5-9%	20	21%
<5%	70	73%
<b>Total</b>	96	



## 5. Checking that the correct digital entities have been included in the fused dataset

(Refer to comScore document "Minimum Reporting Standards (UK).pdf" and Ipsos file "Summary of tagged entities")

Ipsos does not apply any further sample thresholds beyond those which are applied by comScore in the publication of the currency data. Please refer to the comScore document "Minimum Reporting Standards (UK).pdf" for a detailed understanding of comScore criteria for Minimum Reporting Threshold (MRS).

The Ipsos file "Summary of tagged entities" has two worksheets, the first labelled "Tagged but not in Ipsos MML" shows which publisher entities were tagged for PAMCo in November 2017 but which have not been included in the fusion as they were not requested by publishers. Of the 116 entities in this list, only 8 would have reached comScore's MRS in November and could potentially have been published had this been requested. These entities are listed overleaf:

- cheshirelife.co.uk
- cloudpro.co.uk
- cyclingnews.com
- cyclist.co.uk
- decanter.com
- lancashirelife.co.uk
- practicalfishkeeping.co.uk
- whatdigitalcamera.com

The second worksheet labelled "In Ipsos MML" indicates which of the entities are published in the Q4 2017 PAMCo dataset. Those not for publication fall into one of three categories a) missing from comScore's MP file (col F) due to not reaching MRS or not being tagged, b) in MP file but missing from comScore's DemoBreak file (currency targets) due to not reaching MRS (col G) or available but PAMCo or publisher requested us not to publish. This last category applies mainly to digital sections.

Column H shows that there are 82 published digital entities for which Ipsos has modelled the demographics as they were absent from the DemoBreak (currency targets) file provided by comScore.

## Multi-Platform Respondent Level File

*(An outline of the creation process)*

comScore reports digital audience estimates de-duplicated across platforms in the MMX Multi-Platform product which is based on a multi-sourcing approach and on a methodology that integrates census data (content-centric measurement collected via comScore tags and SDKs) and panel data (metered panel measurement). comScore measures digital audiences at the individual platform level first and then de-duplicates audiences across platforms into total digital audience estimates. comScore methodologies work with various input data sources and formats that need to be integrated into a consistent outcome, so the native format of MMX Multi-Platform data suite is not a simple respondent level file but a more complex set of data files. For the PAMCo/Ipsos project purposes, comScore has developed a process that produces the Multi-Platform Respondent Level File from the internal comScore data sources that are used to produce the MMX Multi-Platform data available in comScore's interface.

The Multi-Platform Respondent Level File process joins several dissimilar data streams into a singular combined data set via a specific fusion approach. The net result of the fusion process is a person-based file that provides usage metrics with each individual panellist maintaining a single unchanging projection weight across all platforms and entities.

Before the fusion of desktop and mobile data sets can be done, the input sources, i.e. the underlying data sets, need to be adjusted to be ready for the fusion. As comScore's digital audience estimates for individual platforms are based on the integration of census and panel data (Unified Digital Measurement - UDM) and the outcome of UDM is a data set with multiple weights per individual respondent in each sample, the original platform level data sets need to be recalculated to obtain a single weight per respondent. This is done by a process that reassigns visitation to an entity across panellists with the same demographic profile who show the closest digital behaviour.

Then, the fusion process integrates the adjusted desktop and mobile respondent databases into the Multi-Platform Respondent Level File by pairing of respondents based on the behavioural similarity between visitors of an entity across data sources within a demographic subset of the population (with the same demographic profile). The overlap target is calculated at entity and demographic break level from the original MMX MP data and determines how much UV must be matched between the two databases.

The MP Respondent file creation process is designed to reflect the MMX Multi-Platform data at the entity level within an agreed tolerance range and follows the same logic of digital audience estimates reporting as in MMX MP in comScore's interface. In the MMX MP production process, the audience estimates are processed at the entity level for each entity defined in the Dictionary separately and cross-entity overlaps (that are not defined in the Dictionary) are not available in the MMX MP module and thus are not controlled the MP Respondent file creation process.