Comparing claimed and passive publisher data

Sandra Collins
Andrey Ponomarev
Neil Farrer

Ipsos Connect, UK

Introduction

There is general acceptance amongst media researchers that online activity is best measured through passive means. The many ways of viewing online content and the fleeting nature of many of these visits make it difficult for survey participants to recall the sites that they have visited and when last and how often they visited them.

Since January 2017, Ipsos Connect have been collecting print and online reading data for PAMCo (Publisher Audience Measurement Company) on behalf of the newspaper and magazine publishers in Great Britain. The AMP (Audience Measurement for Publishers) system uses face-to-face interviews, combined with passively collected online data from comScore to measure combined print and digital reach for publishers. A sub-sample of the face-to-face sample have online visits from their computers, smartphones and tablets tracked passively (AMP Panel) and the data from these panellists are used in the process of calculating the size of the overlap between print and online. A paper covering the development of this panel is being presented separately at this symposium (Developing a Quality Passive Panel – Farrer et.al).

Although the AMP Panel is not used to provide absolute levels of print or online reading, the panel provides us with a unique opportunity to gain insight into the nature of differences between claimed online reading of publisher brands and actual online behaviour.

Survey questions about online reading

The AMP project includes a face-to-face CAPI survey of the Great British adult population, based on an annual 35,000 random probability sample.

Questions about online reading are included in the AMP survey as part of a 'brand first' approach. Questions are asked about any viewing of the publisher brand either in print or on screen and then separately for both. The main purpose of including questions about online reading in the face-to-face interview is to help the participant exclude any claims for online reading when answering questions about print.

The following survey questions relating to online reading of specific brands are relevant to our analysis:

1. Identifying any visits to the brand:

   "Have you read or looked at [brand] in print or on screen in the past 12 months?"

2. Identifying whether in print or online:
   "Have you read or looked at [brand] in print or on screen in the past 12 months? Please answer yes or no for in print and on screen"

3. Identifying frequency of online visiting:
   "How often do you look at [brand] on screen, on any device?
   Response codes: every day, most days, about once or twice per week, about once or twice per month, about once every 3 months, less often than every 3 months

Survey questions about device usage are also of interest in our analysis:

1. Establishing personal usage of devices:
   "Please tell me how many of each of these you have in your household and then tell me how many you personally use."
   Response codes included: Computer (excluding work computers), Smartphone, Tablet.

   "…In the last 30 days, have you used any computers or laptops that do not belong to the household…"

If yes> "And how would you best describe ownership of this computer or laptop?"
Response codes include: Owned by your employer or place of work.

2. Establishing primary usage of devices:

"And thinking of the computer that you use, does anyone other than yourself use this computer or laptop?"

If yes> "What percentage of time is spent by people other than yourself on the computer?"

"…please tell me if you are the primary or main user of this phone…"

"…thinking now of the tablet that you use, are you the primary or main user…"

3. Establishing the device type:
For computers: Windows, Apple or MAC, other types
For smartphones: Apple iPhone, Android phone, Blackberry, Windows phone, other types
For tablets: Apple iPad, Android tablet, other types

4. Establishing devices participant used to access the internet in the past 30 days

5. Establishing devices participant used to look at magazine or newspaper content, and frequency of use

"…which of these devices do you use to look at newspaper or magazine content..please answer separately for newspaper content and magazine content"

"…roughly how often do you look at [newspaper / magazine] content on your [computer/smartphone/tablet]..?"
Response codes: several times a day, about once a day, a few times a week, about once a week, a few times a month, about once a month, less often.

Passive data collected

A sub-sample of those interviewed for AMP installed the AMP App[1] on all of the personally owned computers, smartphones and tablets that they use to access the Internet.   Activity is captured for a minimum 4-week period.   For these participants, comScore return the following information to Ipsos:

- Participant ID
- Device/browser used
- Website or App visited
- Number of times visited

As at June 2017 data, 2345 AMP Panellists had installed the AMP App on all of their eligible devices and complete brand information was returned for 74 publisher brands.

---

[1] The AMP App was developed for Ipsos and PAMCo by comScore and allows comScore to track AMP Panellist visits to any publisher website or App that carries a comScore tag.

Claimed visiting versus passive visiting

In this section we examine the differences between claimed survey data from our generic question about looking at newspaper or magazine content on the devices that the participant uses, and site visits appearing in the passive data, from amongst AMP panellists.

At an overall level, more panellists had passive visits to at least one publisher site (68%), than claimed to look at any newspaper or magazine content on any of their devices (62%).

Table 1. Any claimed newspaper/magazine content visits v any passive publisher visits

Total Panellists	Claimed & Passive	Claimed only	Passive only	Neither claimed nor passive
2345	1101	365	486	393
100	46.9%	15.6%	20.7%	17.0%

Base: All panellists

Profile

The younger age groups in our panel are most likely to provide recall data that is in line with their passive behaviour (index of 77) and this declines slightly by age group.   Social Grade ABC1 panellists are more likely to use online, and provide more accurate recall data than Social Grade C2DE.

Table 2. Profile: Any claimed newspaper/magazine content visits v any passive publisher visits

Device type		Any Passive	Claimed & Passive	Index on any Passive	Claimed only	Passive only
Male	100%	80.1	56.6	71	19.9	23.5
Female	100%	82.1	56.3	68	17.9	25.9
Age 15-24	100%	82.6	63.3	77	17.4	19.3
Age 25-44	100%	80.1	59.0	74	19.9	21.1
Age 45-64	100%	81.5	53.1	65	18.5	28.5
Age 65+	100%	82.4	45.8	56	17.6	36.6
Social Grade ABC1	100%	83.4	61.8	74	16.6	21.6
Social Grade C2DE	100%	77.3	46.3	60	22.7	31.0

Base: All panellists with passive or claimed visits (1859)

Devices used

We observed activity on around a quarter of panellist devices that were not claimed as being used to view newspaper or magazine content. Smartphone users accounted for a higher proportion of passive usage, but we did not observe any skews in terms of the ability to recall visiting by device type, with claimed data accounting for around two thirds of passive visits across each device type.

Table 3. Devices used: Any claimed newspaper/magazine content visits v any passive publisher visits

Device type	Total Devices	Any Passive	Claimed & Passive	Index on any Passive	Claimed only	Passive only
Computer	923	60%	38%	63	40%	22%
Tablet	640	56%	35%	63	45%	21%
Smartphone	1534	75%	49%	65	25%	26%

Base: All devices

The claimed usage of devices to view newspaper or magazine content that is not confirmed by the passive data, may in part be attributed to visits made to newspaper or magazine sites that are not comScore-tagged sites.

Claimed versus passive – brands

In this section we compare recall data for visiting individual publisher brands online at least once a month, with activity in the panel data over a 4-week period. To date we have complete data for 74 publisher brands from which we are able to make comparisons.

For ease of analysis we have separated the publisher sites into size and type:

Categorisation of Sites by size

Type of site	Total	Small	Medium	Large
News	18	4	5	9
Special interest	13	12	1	0
Celebrity/gossip	11	10	1	0
Women's Interest	10	7	3	0
Home	7	6	1	0
Automotive	4	4	0	0
Food	3	2	0	1
Retail	3	1	2	0
Men's Interest	3	2	1	0
Entertainment info	2	0	2	0
TOTAL	**74**	**48**	**16**	**10**

Small sites: Net claimed or passive reach <2% of sample
Medium sites: Net claimed or passive reach 2-9% of sample
Large sites: Net claimed or passive reach 10%+ of sample

Unsurprisingly perhaps, the larger sites received more passive visits than were claimed, whereas claimed visits to smaller sites tended to be over-stated.

Table 4. Claims for individual brand visiting at least once a month v passive brand visits over 4 weeks: Size of site

	PASSIVE = CLAIMED	MORE PASSIVE THAN CLAIMED	MORE CLAIMED THAN PASSIVE
Small sites	13%	17%	71%
Medium sites	6%	31%	63%
Large sites	0%	90%	10%

We were interested to note in the 2015 PDRF paper from Experian Marketing Services (Cross Platform Media Measurement: Mobile and Desktop Online Measurement comparisons) the observation that there was "consistent under-reporting for Smartphone and PC behaviors," with PC-users in particular under-reporting.

We see under-reporting for the larger publisher brands in our study, and we observe that the sites in the aforementioned paper were sites of some magnitude. The picture becomes less clear for small or medium sites.

The automotive and food sites had better recall. We can hypothesise that there is a stronger interaction with these types of sites (seeking information for purchasing a car or using recipes on food sites for preparing food).

The majority of news sites fall into the 'large category' with large numbers of visitors who are perhaps more infrequent visitors to the sites, accounting for the number of passive visits exceeding survey claims.

The categories with high claims and low passive data are perhaps those brands with strong brand relevance to the participant which the participant may visit on an infrequent basis and did not visit within the 4-week tracking period.

It should also be borne in mind that recall of visiting publisher brands is based on behaviour at the time of interview, and the passive data is captured in the 4-8 weeks after the time of the interview. This too can contribute to differences in the claimed and passive data.

Table 5. Claims for individual brand visiting at least once a month v passive brand visits over 4 weeks: Type of site

|	PASSIVE = CLAIMED	MORE PASSIVE THAN CLAIMED	MORE CLAIMED THAN PASSIVE
News	0%	78%	22%
Special interest	8%	0%	92%
Celebrity/gossip	0%	9%	91%
Women's Interest	10%	10%	80%
Home	14%	29%	57%
Automotive	75%	25%	0%
Food	33%	67%	0%
Retail	0%	0%	100%
Men's Interest	0%	0%	100%
Entertainment info	0%	50%	50%

Frequency of reading

Whilst there are differences between the levels of passive reach and the recall data, we observe a correlation between the number of those visiting websites and the claimed frequency of visiting.

Table 6. Average monthly visits in passive data categorised by average claimed frequency of visiting publisher brands

Claimed frequency of visiting	Average monthly visits from passive visitors (1+ visits during a month)	Average monthly visits from total sample (0+ visits during a month)
Every day	21.8	7.0
Most days	8.3	2.6
About once or twice per week	4.2	1.2
About once or twice per month	3.3	0.5
About once every 3 months	3.2	0.3
Less often than every 3 months	2.1	0.2
Do not read	2.4	0.1
Total visits	**7886**	**192290**

As seen from the chart below, participants who have claimed more frequent visitation of the website are also more likely to be captured as website users by passive measurement.

Figure 1.
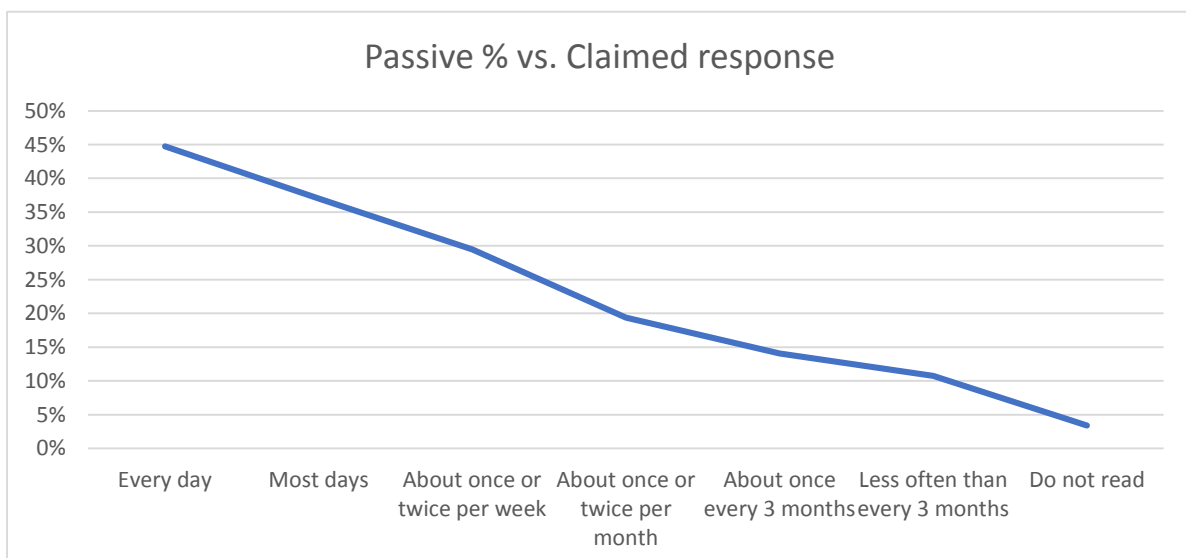
We observe in the passive data, that the more frequent visitors of a site are also those most likely to have claimed any visits to the site. As seen in table 7, there is a threshold of around 15 visits a month, after which the majority of participants start correctly claiming usage of the website.

Table 7. Average monthly visits in passive data v claimed visits to the site

No. of passive visits	0 passive visits	1 to 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	30+
Did not claim	97%	83%	58%	52%	39%	23%	27%	23%
Claimed	3%	17%	42%	48%	61%	77%	73%	77%
	184404	**7218**	**309**	**100**	**71**	**62**	**82**	**44**

These findings are of particular importance as they prove the theory that the more frequent visitors, or those with stronger bonds with a press or magazine brand, are more likely to claim website visitation. This has a few important implications for recall data:

1. Infrequent visitors of sites are strongly under-represented
2. Average frequency is over-estimated
3. The profiles of claimed data users would be skewed toward a website's core, most engaged audiences.

Claimed versus passive – the role of print reading

We analysed print readers and non -print readers in the recall and passive data, to understand the relationship between print and digital reading in the data.

We see that passive visitors are twice as likely to claim visitation if they are print readers (194 index) compared to non-readers. This relationship is even stronger amongst non-passive users, with print readers being 7.5 times more likely than non-readers to claim visitation.

Table 8. Print readers and non-print readers analysed by claimed and passive digital reading

	Non-print reader		Print reader		Index Print readers/non-print readers	Index Print readers/non-print readers
	Not in Passive data	In Passive data	Not in Passive data	In Passive data	Not in Passive data	In Passive data
Did not claim online brand	97.8%	82.3%	83.2%	65.7%		
Claimed online brand	2.2%	17.7%	16.8%	34.3%	747	194
Visits	178729	6812	5657	1074		

Aggregated percentages of individual brands in recall data v passive data

It should be noted that recall data were collected using a brand-centric design. Participants are first asked about looking at the brand in the past 12 months, then on which platform (print or on screen) and finally frequency of visiting digital and recency and frequency of reading print.

This brand-centric design, whilst minimising the respondent load and potentially more accurate for providing individual reach of each of the platforms than asking about platforms separately, might over-represent duplication between print and digital. We expect that there is an element of participant "conditioning" through asking about these platforms together.

Summary

As already observed by a number of studies, we see that the levels of passive data exceed those of recall, but this is not true for all brands. There is some evidence that the younger age groups and those in higher Social Grades are better able to recall digital reading, but there are no strong skews in recall amongst types of devices used according to the AMP data.

Larger sites have a greater proportion of participants that are not able to recall visits to their sites. There is a closer match between passive and recall for particular categories of site, where we hypothesise that there is a greater level of interaction.

We see that recall data is strongly skewed to those participants who are frequent visitors of the site. We also observe that that within our Brand First interview, there is a strong tendency for print readers of the brand to over-claim online reading of the brand.

For Ipsos Connect and PAMCo this underscores the importance of using passive sources to measure online reading and has implications for the way in which recall data are used as part of the fusion process with external digital data.

References

Developing a Quality Passive Digital Panel ((N. Farrer, M Canayaz (Ipsos Connect) V. Smeka, J. Ranta (comScore)) Symposium Paper 2017

Cross Platform Measurement: Mobile and Desktop Online Measurement Comparison. (Pellegrini, P. A., Bradley, D., Swift, K., and T. Boals). PDRF paper 2015