Appendix O: measurement issues in digital advertising

Summary

1. For there to be effective competition between buyers and suppliers of digital advertising, advertisers need to be able to assess and evaluate the quality and effectiveness of the digital advertising inventory that they are buying. If they can do this, then they can make an informed choice between the different types of advertising that are available to them and so force firms to compete for their business.

2. Assessing and evaluating the quality of digital advertising is a process which involves a number of different stages:

    (a) **Verification:** checking the viewability of the advertising, the context in which it was displayed, and identifying the potential for ad fraud. Was the advert displayed on a webpage in such a way that consumers could actually view it?

    (b) **Attribution:** tracking what actions the consumer took after being exposed to the advert. For instance, did the consumer click through to the advertiser's website and buy the product after exposure to an advert?

    (c) **Measuring effectiveness:** did the advertising meet the campaign objectives the advertiser had set eg did the advertising produce an incremental uplift in sales?

3. Advertising campaigns have the characteristics of 'experience goods'. That is, it is only possible to assess the quality of a campaign after it has been purchased and the advertising served to consumers. Key to being able to assess and evaluate the quality of digital advertising is the ability of advertisers (or their agents) to access user-level data to be able to carry out verification, attribution and measurement activities.

4. In general terms, firms with market power will have an incentive to use their position to degrade the price-quality trade-off for the product they supply if that will reduce costs. Part of that degradation in quality can involve restricting the ability of their customers to properly evaluate the quality of the product they are buying so that they are not in a position to make accurate comparisons with the products offered by other firms.

5. In the case of digital advertising, there are concerns that firms with market power have the ability to obstruct or place restrictions on advertisers

accessing and assessing the data they need to carry out a proper, independent evaluation of the quality of the advertising inventory they have purchased. Our concern is that if advertisers are not able independently to assess the relative merits of advertising across different platforms and are forced to rely on the largest platforms' own measurement tools, then they could be over-paying for the advertising inventory supplied by those platforms and mis-allocating their advertising expenditure relative to other sources of supply.

6. Furthermore, if advertisers are forced to rely on information and metrics provided by those platforms, this can make it more difficult for other platforms to demonstrate that they can offer a competitive alternative and so runs the risk of undermining effective competition.

Verification

7. As indicated above, verification is an important first step in measuring the quality and effectiveness of advertising. It also plays an important role in identifying and combatting ad fraud, so that the ability to 'verify' advertising is important in promoting trust and confidence in digital advertising overall.

8. Historically, there have been a number of mis-reporting issues in relation to Facebook and concerns about brand safety on Google owned properties. There are also ongoing concerns about the potential for ad fraud. This in turn has raised concerns about trust and transparency.

9. It is very important that advertisers should be able to verify that the inventory that they purchased has been delivered as contracted for, in particular in respect of viewability and brand safety, and that they have not been subject to ad fraud. In addition, enabling third-party verification allows the performance of platforms to be properly evaluated and compared and this should help stimulate competition.

10. Although Google and Facebook do allow third-party verification of their own inventory, they place restrictions on the ability of advertisers to carry out their own independent verification. That is, advertisers have to rely on data that has been collated by Google and Facebook. This is in contrast to the situation on other platforms.

Attribution

11. It is rarely the case that the consumer 'journey' from seeing an advert to making a purchase is a simple, linear one-step process. An advertising

campaign may run for a number of weeks and involve different advertising channels. This means that a consumer may be exposed to an advert several times across a number of different marketing channels.

12.   A key attraction of digital advertising is the ability to track an individual consumer and to log the actions that consumers take after they have been exposed to an advert. Before the advent of digital advertising, following the 'consumer journey' from advert to conversion was something that was done at a more macro-level and was a more drawn-out process.

13.   Attribution is the process of assigning 'credit' for a successful marketing outcome (eg a customer making a purchase, visiting a website) to the marketing interactions that occurred prior to that outcome. Attribution is thus important for the design of advertising campaigns in terms of deciding the initial allocation of an advertising budget across different marketing channels to achieve a set of campaign objectives. It can then also be important in terms of allowing advertisers to adapt campaign performance in real time.

14.   To be done accurately, attribution relies on the ability to collect data on a consistent basis about an individual user across different devices and across different platforms. Although many parties do have access to data about consumer behaviour eg through access to first or third-party data, the process of attribution is challenging. Google and Facebook have a competitive advantage in terms of being able to carry out attribution accurately for campaigns that advertisers run, at least in part, on their own 'walled garden' platforms. At the same time, Google and Facebook's tracking solutions are widely distributed across many other websites.

15.   In this context, this superior access to data in terms of quantity and/or quality of campaign data on Google and Facebook's properties constitutes a barrier to entry and expansion, albeit not insurmountable, for smaller rivals in the provision of personalised advertising. Competition concerns are, however, increased where a platform with market power also takes actions that make it more difficult for third parties to implement their own attribution solutions and so increases the reliance of advertisers and media agencies on the analytical products and services products offered by the platform.

16.   Concerns have been expressed that actions by Google in particular have made attribution by third parties more difficult, for instance, by no longer sharing DoubleClick user IDs. Google's proposals to block third-party cookies in its Chrome browser in future are likely to make this third-party attribution still more difficult because it will constrain the ability of independent analytics providers to track consumer behaviour across different websites.

17.   We note that Facebook and Google (and others) have made available services such as data 'clean rooms' to advertisers to allow them to carry out the analysis of campaign performance in a controlled environment. We recognise that – over time – this could offer a way forward. However, feedback from media buying agencies indicates that this is still a nascent area. At present the take-up of such services by advertisers appears to be very limited and there are concerns that it is not possible to carry out analysis at the level of individual users. The fact that advertisers are not able to extract data from these environments is also considered to be a constraint on their usage.

Measuring effectiveness

18.   Academic research has pointed to important concerns about standard, observational approaches to measuring the effectiveness of digital advertising. Specifically, the targeted nature of many forms of digital advertising – which makes it so attractive to advertisers – is likely to mean that standard observational approaches produce biased[1] estimates of the effectiveness of that advertising. In the extreme, some research has suggested that the purported effectiveness of digital advertising could be completely illusory, ie it has no effect at all.

19.   If advertisers cannot measure the effectiveness of their expenditure on digital advertising accurately and robustly, then this could lead to a significant misallocation of expenditure between different advertising media. Such results would also tend to undermine trust and confidence in digital advertising.

20.   In response to these concerns, experimental methods have been developed to address these measurement concerns. Properly designed and implemented Randomised Control Trials ('RCTs') have the ability to produce robust estimates of the effectiveness of digital advertising. Studies using RCTs have established that digital advertising is capable of having a positive causal effect on outcomes such as purchases and website visits.

21.   Discussions and information requests to media agencies indicate that agencies and advertisers are aware of the methodological challenges around the measurement of the incremental impact of digital advertising and are aware that experimental approaches offer a more robust way to do this than observational approaches. In addition, both Google and Facebook have been

---

[1] Here we use the term 'biased' in a specific, statistical sense. That is, estimates of the effectiveness of advertising will differ systematically from the actual effectiveness.

involved in the development of the academic research and offer advertisers the tools to carry out different forms of RCTs.

22.   There is some evidence to suggest that advertisers in the UK are moving towards making use of a range of experimental approaches to evaluate campaign effectiveness. However, this practice does not appear to be as widespread as might be expected, given the sums of money invested in digital advertising. Media agencies still report that factors such as cost and complexity constrain the use of RCTs on a systematic basis: advertisers need to commit to funding the measurement of the effectiveness of advertising campaigns. At the same time, there do appear to be some practical constraints on their deployment on a widespread basis. For instance, it may be difficult to implement RCTs across all advertising campaigns and RCTs may not be available across all advertising formats.

23.   We note that advertisers typically do not rely on just one measurement approach to evaluate campaign effectiveness. Instead they will make use of a range of different metrics (including metrics that focus on the delivery of an advertising campaign) and exercise their judgement depending on the nature of the advertising campaign.

24.   Overall, while techniques have been developed to overcome key measurement issues and Google and Facebook do make available the tools to assess the effectiveness of advertising on their platforms. However, our analysis also points to the importance of advertisers and agencies being able to evaluate the effectiveness of digital advertising and to continue to engage with platforms about the tools and metrics that they need to access in order to be able to do this and to do this on an independent basis as necessary.

Introduction

The role of advertising

25.   Conventionally advertising can be thought of as having a number of different purposes – persuasive, informative and complementary – and the impact on consumer welfare will depend on the purpose. For example, where advertising improves the information available to consumers it will also improve consumers' knowledge and understanding of prices, product quality etc, helping to grow demand in a market and so improve consumer welfare.

26.   In contrast, where advertising is persuasive and is primarily focused on shifting demand between products, then it can have a negative impact on consumer welfare.

27. All advertising involves a degree of ex ante targeting to a greater or less extent. For instance, advertisers purchasing TV or radio advertising will look to target certain audiences based on broad demographic characteristics (eg Adult, Male, 16-34 years old). Similarly, advertisers using outdoor advertising will look to buy space on billboards in certain areas based on information about the population in or passing through that area.

28. However, in the case of personalised digital advertising, advertisers potentially have access to a large volume of historic user-level data on everything from the interests of potential customers, the devices they use and their location at any point in time. This provides the ability to target audiences more precisely and so reduce the 'wastage' that is a feature of traditional advertising media. The use of cookies and other methods should also mean that advertisers are better able to track a consumer's journey across the internet in terms of the adverts that they see and the actions that they take after having seen those adverts.

29. In general terms, better targeting of advertising – whether contextual or personalised – could also represent a benefit to consumers. Targeting should mean that advertising can provide useful information to a consumer about the product or service they are interested in.

30. However, consumers can also regard too high a level of personalisation as a nuisance. Consumers can be wary of being tracked too closely by firms and this can prompt concerns about the invasiveness of advertising from a privacy point of view and, in turn, provoke reactance on the part of the consumer.

31. In order to avoid eroding the value of personalised advertising in the first place, both media owners and advertisers should have an interest in managing the targeting of advertising to avoid consumers dis-engaging. We recognise that there could be a 'free-rider' issue in that individual advertisers may not care too much about the impact of their adverts on consumers' overall engagement in the short-run but media owners will have a direct interest in maintaining user engagement and keeping users on their sites.

Measuring the quality of digital advertising

32. Advertising campaigns have the characteristics of an 'experience good' in that it is only possible to evaluate its quality and effectiveness after an advert has been shown to an audience. In fact, advertising campaigns could be regarded as 'credence' goods in that it can be difficult to and costly for advertiser to

measure the incremental effectiveness of an advertising campaign even after it has been run.[2]

33.   It is important – for the effective functioning of competition in digital advertising markets – for advertisers to be able to properly measure and assess the 'quality' of advertising inventory and so to make informed choices about the different types of advertising that they buy. If advertisers are not able to assess the quality and measure the effectiveness of the digital advertising they are buying, then they are not able to impose a competitive constraint on the supply side of the market.

34.   Although one of the attractions of personalised digital advertising is the volume of user-level data available, at the same time the complexity of digital advertising can make it difficult for advertisers to monitor and evaluate the effectiveness of advertising. In particular, many advertisers will be small and micro-businesses or even individuals, which could also mean that their understanding of the operation of the market is more limited.[3]

35.   The risk of ad fraud (eg advertisers paying for impressions that cannot be seen or clicks that have been made by 'bots' rather than humans) also further complicates the process of monitoring and evaluating the quality of digital advertising.

36.   The measurement of quality is a process which involves evaluating a number of different aspects of digital advertising:

        (a) **Verification:** checking the viewability of the advert, the context in which it was displayed and identifying the potential for ad fraud. For instance, was the advert displayed on a webpage in such a way that consumers could actually see it?

        (b) **Attribution:** tracking what actions the consumer took after being exposed to the advert. For instance, did the consumer click through to the advertiser's website and buy the product after exposure to an advert?

---

[2] An individual advert would be a 'credence good' in that it would be extremely difficult and costly to measure the effectiveness of an individual advert and the results are likely to be imperfect. As discussed in this Appendix, advertisers are in a position to estimate the incremental impact of ads on conversions. However, as it also noted in this Appendix, advertisers do not choose to measure the incremental impact of ads on a regular basis. As a result, advertising campaigns could be considered to be experience or credence goods.
[3] As set out in more detail in Appendix N, as part of our investigation into the behaviour and characteristics of advertisers, the CMA commissioned an independent research agency, Jigsaw, to collect views from a range of smaller advertisers and summarised these responses. Some of these advertisers did indicate that their choice of advertising platform was driven by a lack of knowledge of effective alternatives. Others stated that they viewed their current situation of using only one platform as 'good enough', with acceptable ROI and perceived good value for money.

(c) **Effectiveness:** did it meet the campaign objectives the advertiser had set, eg did the advertising produce an incremental uplift in sales?

37. A number of concerns have been expressed about the different aspects of how digital advertising is monitored, measured and evaluated. These concerns centre on access to and control over the data used in the various stages of the process of measuring quality.

Verification

Introduction: viewability, brand safety and ad fraud

38. Verification encompasses a number of different aspects of the quality of digital advertising: viewability, brand safety and safeguarding against ad fraud.

39. The techniques that are used to verify viewability are also used to safeguard brand safety and protect against ad fraud across different digital channels, ie desktop, mobile web and mobile app.

40. Verification technology is primarily used by advertisers and DSPs but can also be deployed by publishers to help them better monetize their inventory.

Viewability

41. When an advertiser receives a bid request from a publisher, the bid request will include a description of the advertising inventory (or impression) that is being offered for sale at that point in time. For instance, the bid request will include details of the position of the ad on the webpage, whether the ad is to be viewed vertically or horizontally, the device the consumer is using to access the website (eg laptop versus mobile handset), etc.

42. Verification of viewability involves the authentication of the placing of an advert on the website, how much of it was viewable and how long it was viewable for.[4] The fact that an advert was viewable does not guarantee that an advert was seen by a consumer, only that the advert had the opportunity to be seen. Viewability levels are taken into account when determining what advertisers pay for impressions delivered. Viewability levels are also a useful metric for the owners of websites in that it provides feedback to help them optimise the layout of the website and page experiences to increase the viewability levels of their ads.

---

[4] For instance, the IAB defines viewability in terms of a 50% of the ad unit being in view for a minimum of 1 second for standard ad formats and a minimum of 2 consecutive seconds for video ads. See IAB Quick Q&A Viewability.

Brand safety

43.   Viewability is closely linked to the issue of brand safety, ie ensuring that an advert does not appear alongside inappropriate content or content that is not in keeping with an advertiser's brand value.

44.   Verification in relation to brand safety can involve both pre- and post-bid verification. For 'pre-bid' verification, advertisers can set verification thresholds in advance through the DSPs they use. Before each bid, a DSP will check the page-level URL or the mobile app sending out the bid request against criteria set by the client before determining whether to submit a bid. If a bid request fails the pre-bid checks (eg the bid request comes from a website that is not on an approved 'whitelist' set by the advertiser), then the DSP will not submit a bid for that impression. If post-bid verification established that an advert was served alongside inappropriate content or content not in keeping with the advertiser's brand values, then typically the advertiser would not be charged for that impression.

Ad fraud

45.   If fraud is prevalent in digital advertising markets, then that will undermine the functioning of those markets. If advertisers cannot be sure that the advertising inventory they are buying is authentic or that the agents they are trading with are legitimate, then that lack of transparency will lead to a lack of trust in digital advertising. As indicated above, because of the experience / credence good nature of digital advertising campaigns, the usual incentives on firms to act in a way to preserve their reputations may not be sufficient to address this concern.

46.   Verification can also involve checking for fraudulent activity. For instance, it is possible to verify IP addresses and check if they have been infected, are proxies, or are simply data centres which are generating bot traffic. Again, however, verification will not always be perfect.

47.   There are a wide range of estimates of the overall scale of fraud in relation to online advertising. However, a number of stakeholders referenced the most recent report[5] by White Ops and the American Association of National Advertisers (ANA), which estimated that the global scale of losses from ad fraud would fall from $6.5bn in 2017 to $5.8 bn in 2019 despite estimates that digital ad spending had increased by just over 25% over the same time

---

[5] Bot Baseline Report: Fraud in Digital Advertising (May 2019) The report by White Ops and the American Association of National Advertisers (ANA) involved 50 ANA member companies and tracked sophisticated invalid traffic (SIVT) across 130,000 ad placements and 27 billion ad impressions over the course of two months.

period. The total amount spent on digital advertising globally in 2019 was estimated to be around $267bn in 2019 which suggests a fraud rate of just over 2%.[6] The report made an important distinction between attempted fraud – where it estimated that fraudulent attempts accounted for 20-35% of all ad impressions – and the amount of fraud that actually got through. It found three main factors which contributed to this projected decline in ad fraud:

(a) The use of ads.txt to help publishers create lists of authorized media sellers.[7] This had worked to reduce desktop spoofing to the lowest levels recorded in the history of the report.

(b) It had become more expensive and less efficient to buy sophisticated bot traffic. Efforts by the Trustworthy Accountability Group (TAG) and their Certified Against Fraud program, coupled with groups working together to dismantle botnets, had drastically reduced both the supply and the demand for traffic from vendors that are caught selling bot traffic.

(c) More digital advertising was being sold through platforms that had built-in fraud prevention measures.

48. Responses from advertisers and agencies also appear to indicate that the level of ad fraud in the UK is estimated to be relatively low. For instance, a number of responses indicated that the scale of ad fraud was estimated to be between 1-2% of impressions although, by its very nature, successful fraud would tend to be under-reported. There is a perception that ad fraud is more of a risk in relation to the open display market because of a long tail of smaller publisher sites.

49. Respondents were also aware of the main industry initiatives by the IAB, JICWEBs and others aimed at addressing brand safety and fraud issues, and many respondents were also active participants in those industry initiatives. At the same time, some agencies also reported that such initiatives were not yet fully mature and there was a general recognition that there was more work still to be done to combat ad fraud.[8]

50. Overall, the verification of digital advertising is an important first step in measuring the return on investment (RoI) of an advertising campaign. Ad

---

[6] Estimates of total global spend on digital advertising taken from: Statista: Global Internet Advertising Expenditure by Type.

[7] Ads.txt stands for Authorised Digital Sellers. It is a method that allows publishers to publicly declare the companies they authorize to sell their digital inventory. It can thus provide advertisers with more confidence that they are buying authentic inventory. The text file can be updated to provide for flexibility over time.

[8] For instance, in its response to our interim report, JICWEBS referred to the fact that it was currently facilitating a pilot into blockchain / Distributed Ledger Technology (DLT) over the course of 2019/ 2020 to understand more about how this technology could bring greater transparency to digital advertising transactions. JICWEBS Response to Interim Report

verification checks that advertising has been displayed on the right websites, on the right part of those websites and was available to be seen by the correct audiences. Verification does not, however, prevent against ads from being displayed in the wrong place. By the time the ad is served to a website, the auction for that impression has already taken place. Rather, verification should mean that advertisers do not pay for adverts that are incorrectly displayed or are not visible to a consumer. If there are issues around viewability, brand safety and ad fraud, then that will detract from the effectiveness of the advertising and its RoI.

How is verification carried out?

51.   When an advert is served to the publisher's webpage, it can include a verification 'tag'. The verification tag will instruct the user's browser to connect back to a server to allow the collection of basic data about the URL and viewability.

52.   For large advertisers, the ad verification process is something that is typically carried out by a third-party ad verification provider who is contracted for by the advertiser or their media agency. Alternatively, a verification provider might work with a DSP and the verification process can be integrated into a DSP.

53.   In the case of mobile apps, measuring viewability requires SDKs ('Software Development Kits') to be installed in the app by the developer. If a publisher decides not to integrate a vendor's SDK into their app, then the vendor and client will struggle to measure viewability.

54.   An SDK enables third parties to integrate tools into the apps. Historically, there was a need for an SDK for each vendor, but each SDK would increase the size of the app and increase latency. The IAB has introduced an open measurement SDK which provides a single standardised SDK for viewability rather than requiring one per vendor.[9]

55.   The tag collects data to be able to analyse the content of the publisher's page and to report information about the placement, viewability etc of the advert back to the advertiser or verification service provider.

56.   Verification tags can also help with detecting geographic targeting and frequency capping in advertising campaigns.

---

[9] IAB Open Measurement SDK.

Historic mis-reporting of data

57.   It is possible that a reason that there are concerns about trust and transparency in respect of digital advertising is that there have, historically, been issues to do with the mis-reporting of data by Facebook and concerns about brand safety in relation to Google.

Facebook

58.   Over the course of September to December 2016, there were a series of reports about Facebook's mis-reporting of important advertising performance metrics. In a number of cases, the issue was identified by external parties – in other cases, it was Facebook that revealed the measurement issue.

59.   For instance, in September 2016, Facebook acknowledged that it had over-reported the average 'watch time' metric for videos. It was reported that the extent of the overstatement was estimated to be between 60-80% for around two years. A lawsuit over this matter was reportedly settled in 2019.[10]

60.   Over the period September to December 2016, it was also reported that Facebook was found to have been: inflating the number of organic visits to brand posts; over-reporting by 7-8% the average length of time people spent reading Instant Articles; over-stating video ad completion rates; overreporting the number of clicks it sent to advertiser websites; over-reporting the number of Likes for live videos; and inflating the number of times people shared links of posts on Facebook.[11]

61.   Following on from this example, it was then reported in 2017 that videos that were loaded on Facebook's mobile site continued to play after they were scrolled out of view, leading Facebook to charge advertisers for the background views. Facebook is reported to have refunded advertisers in full.[12]

62.   Finally, in September 2017, it was reported that media agencies became aware that Facebook materials claimed to reach more people in the U.S. than Census data showed existed. The claims were made in Facebook's self-serve interface that businesses use to evaluate and purchase Facebook advertising. These claims are now the subject of a class action complaint.[13]

---

[10] WSJ: Facebook reaches proposed settlement in video measurement lawsuit.
[11] Marketingland.com, 'FAQ: Everything Facebook has admitted about its measurement errors.'
[12] Marketingland.com, 'FAQ: Everything Facebook has admitted about its measurement errors.'
[13] Marketingland.com, 'FAQ: Everything Facebook has admitted about its measurement errors.'

Google

63.   In the case of Google, in 2017, it was reported that a number of large
       advertisers were reported as having removed their advertising from YouTube
       in response to concerns that their adverts were being shown alongside videos
       promoting extremist and terror groups.[14]

64.   These concerns appear to have continued through to the end of 2018, with a
       number of advertisers reporting that they had suspended advertising with
       YouTube because of brand safety concerns.

Access to verification data

65.   In our interim report[15] we reported that some advertisers had raised concerns
       about restrictions on the third-party verification of advertising on inventory
       owned and operated by Google and Facebook. Although it was the case that
       both Google and Facebook worked with 'approved' third-party verification
       providers, they restricted access to the detailed consumer level data needed
       for verification of advertising on the advertising inventory that they owned and
       operated. Other platforms have told us that they allow advertisers to use
       tracking tags for third-party verification of impressions served on their
       advertising inventory.

66.   Without access to the underlying raw data and the ability to have full
       independent verification, there was a perception on the part of advertisers and
       agencies that Google and Facebook were able, in effect, to 'mark their own
       homework' in respect of the effectiveness of their own advertising inventory.

67.   Both Facebook and Google had argued that the way that they compiled data
       for the purposes of verification – for instance, on the viewability of impressions
       on their inventory – met industry standards and was subject to external audit.

68.   In our interim report, as part of a series of potential interventions to improve
       transparency in relation to digital advertising, we put forward the suggestion
       that there should be a requirement on SMS platforms to provide sufficient
       data for effective ad verification purposes.[16]

---

[14] In 2017, it was reported that a number of large UK advertisers had pulled their advertising from YouTube in
response to concerns that their adverts were being shown alongside videos promoting extremist and terror
groups. Source: The Independent. Banks join queue of advertisers ditching Google over extremist YouTube
videos.
[15] CMA Digital Platforms and Online Advertising: Market Study Interim Report. Paragraph 5.124-5.125.
[16] CMA Digital Platforms and Online Advertising: Market Study Interim Report. Paragraph 6.178.

Responses our interim report

69.   In its response to our interim report consultation, Facebook re-iterated its position that it did work with third-party verification providers, who carry out regular checks on Facebook's ad viewability and other attention metrics. Facebook also reiterated its point that its measurement of impressions for ads on Facebook News Feed and Instagram was accredited by the Media Rating Council – the organisation that sets industry measurement standards for television and radio – and that its systems were subject to audit.

70.   Other stakeholders supported our finding that restricting data flows to independent ad tech service providers was making it increasingly difficult to provide independent analytics, measurement and ad verification services. For example, Oracle Moat argued that Google should make raw data available to advertisers or publishers such that they can reach their own conclusions as to the effectiveness of their ad campaigns. In addition, it argued that Google should not prevent third parties from collecting data directly from its owned properties, such that third parties were not simply measuring Google's curated data but could provide independent measurement and verification services.

71.   Oracle Moat also drew attention to the announcement by Google of the concept of a 'privacy sandbox' for the Chrome browser. It argued that the proposal would have an especially detrimental impact on third-party verification providers because it would restrict their ability to detect bot traffic.[17]

Our assessment on verification

72.   For there to be effective competition in a market, it is important that consumers are able to discipline the supply side of the market eg by switching to other providers if they are dissatisfied with the product or service offered by their existing supplier.

73.   In order to do this, consumers ideally need to have access to information about the quality of the services or products they are purchasing and to have confidence in the veracity of that information. In the case of advertising campaigns, the quality of the advertising inventory – as an experience / credence good – can only really be evaluated after it has been purchased. This means that an important first step in this assessment and evaluation process is for advertisers to have the ability to confirm that the advertising they have contracted for has been delivered as agreed and contracted for.

---

[17] Oracle Response to CMA Interim Report.

74.   However, advertisers also need to be able to switch to other suppliers if they do not have confidence in the quality of the product they are being supplied with. If advertisers do not have much choice but to use SMS platforms then even if they have concerns about verification, they are not in a position to do much about that. At the very least, advertisers should be able to carry out verification for themselves and should not have to rely on verification being provided by the owner of the advertising inventory.

75.   By not allowing full independent verification of their own inventory, Facebook and Google have introduced a degree of opacity into the buying and selling of advertising of their own inventory. It is clear that the buying and selling of display advertising is a complex process so that introducing additional restrictions on access to data adds to the 'friction' in terms of evaluating market outcomes.

76.   We do recognise that there could be potential privacy concerns in relation to providing access to certain user-level data for verification purposes eg invalid traffic and frequency capping. However, we have been told that as the information necessary to carry out verification for viewability and brand safety does not necessarily need to rely on personal data, it is capable of being carried out in a privacy-friendly way.

77.   It is not clear to us that the data involved in verification of viewability and brand safety necessarily needs to involve personal data. For instance, we understand that – for viewability – verification involves determining whether the ad was served, whether the ad appeared on the screen, how much of the ad appeared on the screen, for how long the ad appeared on the screen, whether the ad played (if it was a video), for how long it played (if it was a video) and whether the sound was on (if it was a video). It should be possible to put in place contractual arrangements to facilitate the sharing of data necessary for independent verification of viewability and brand safety.

78.   In terms of ad fraud, here too a lack of transparency can undermine the functioning of a market. That is, if advertisers cannot be sure that the advertising inventory they are buying is authentic or that the agents they are trading with are legitimate, then that lack of transparency would undermine trust in an industry. However, the responses to our information requests and our interim report indicate that ad fraud is generally recognised as an industry-wide challenge and an issue that requires industry-wide solutions to address it. As such, it is not clear that intervention is required to address this issue at this stage.

79. Large platforms – such as Google and Facebook – do appear to be engaging with industry initiatives to combat ad fraud eg in terms of seeking industry accreditation for their fraud prevention processes and systems.

Attribution

Introduction

80. Attribution is the process that identifies a set of consumers' actions across websites and devices that contribute in some way to a desired advertising outcome.

81. The process of attribution typically involves the matching of data of consumers' exposure to advertising across different marketing channels with data on their subsequent actions. The consumers' actions that are most often monitored are purchasing behaviour, but 'conversions' can be defined more broadly depending on the campaign objectives, eg to include visiting a specific website or registering for more information on a product.[18]

82. Advertising campaigns typically run over several weeks and may involve a number of different creative treatments and different marketing channels. This means that the consumer that counts as a 'conversion' may have seen the same advert on different websites or have seen different forms of the advert over the course of the same advertising campaign. In the 'Consumer Journey' diagram set out below, a consumer may have initially heard about a good or service through 'word of mouth'. They may then have seen an advert for this product while visiting a social media platform, looked up some online product reviews as part of carrying out their own research and then searched for different outlets that stocked the product before finally making a purchase.

Figure O.1: the consumer journey

Source: CMA

---

[18] Attribution tends to be associated most directly with advertising campaigns that have a specific, measurable objective such as conversions as opposed to branding campaigns that are focused on raising awareness of a brand.

83.   The process of attribution involves not just tracking the different steps taken by the consumer but also deciding how much credit or weight to give to those different steps in determining the final outcome.

84.   An additional complication is that a consumer could have been exposed to one set of advertising on a mobile device but then subsequently made the purchase online on a different device such as a tablet or a desktop computer, or indeed may have completed the purchase in-store. To measure conversions, advertisers need to be able to track consumer actions across devices and – to some extent – offline as well.

85.   The results of the attribution analysis are typically used to formulate and evaluate marketing campaigns eg to determine how best to allocate advertising resources to achieve particular campaign objectives. However, it is important to stress that attribution on its own does not establish causality. That is, attribution captures the correlation between a consumer's exposure to particular marketing channels and a successful outcome: it does not establish that the final conversion was caused by the consumer's exposure to the marketing. Further analysis would be needed to establish the incremental impact of advertising.

86.   Nevertheless, advertisers do use different attribution methodologies in different campaigns to inform their analysis about which marketing channels are the most important in terms of prompting the final conversion,

Conversion tracking

87.   Tracking user behaviour for attribution and ad measurement purposes will typically involve the use of cookies or other tracking technologies such as pixels or tags.

88.   As described in more detail in Appendices F and G, cookies are small pieces of code which are facilitate the collection of information about the user's visit to a website.

89.   When a user visits a website or mobile app for the first time, the host website or mobile app will 'set' a number of cookies on the user's browser or mobile device to collect data about the user and how they interact with the website.[19] This constitutes first party data from the perspective of the website owner. Cookies from other firms ('third-party' cookies) may also be loaded on to the user's browser (with the appropriate consent from the user).

---

[19] Tags and pixels, also known as web beacons, are similar to cookies but are embedded in pictures and images.

90.   A cookie will assign a unique ID to that user or, more precisely, to that user's browser or mobile device: each third-party cookie will allocate their own ID to that user.[20] According to the type of cookie, that cookies will collect information about that session. The data from that session will then be associated with that user's ID. A user can be tracked across a range of different websites by a range of different organisations

91.   In the case of conversion tracking, when a user clicks on an ad, a cookie is placed on the user's browser or mobile device. If the user then completes a defined action (makes a purchase, registers on the website, makes a phone call, installs an app, etc) that action is recorded as a conversion.[21]

Attribution methodologies[22]

92.   Many standard attribution methodologies are based on straight-forward rules.

Last touch attribution

93.   A very common attribution model is based on 'last touch' attribution model. This means that 100% of the 'value' for a conversion is given to the last advert or marketing message that the user was exposed to before they clicked through to the advertiser's website.

94.   Last touch attribution is a common feature of many analytics tools because it is a simple concept both to implement and to understand. For instance, Google Analytics uses a permutation of this approach – 'last, non-direct click' attribution – in which direct traffic to a website is ignored and 100% of the value for the conversion is given to the last channel that the customer clicked through from before buying or converting.[23]

95.   Moving away from last touch attribution, an alternative approach is 'first touch' attribution. As the name suggests, under this approach 100% of the value is allocated to the first impression that the consumer interacted with.

---

[20] Third-party cookies can be set by range of different companies. For instance, as well as setting its own (first party) cookies from its website, the Guardian newspaper partners with a wide range of third-parties which set their own cookies on the browser of a user who visits the guardian.com These include 'performance' cookies from Comscore and Google Analytics as well as 'advertising' cookies on behalf of advertising platforms including Google, Facebook , Twitter and Bing together with various ad tech providers. Each of these partners will then collect data on a user each time they visit the Guardian website. https://www.theguardian.com/info/cookies
[21] For a description of how this is managed by Google, see How Google Ads tracks website conversions.
[22] See IAB Attribution Hub for more detail on approaches to attribution.
[23] See Google's About the default MCF attribution models. We note that it can be argued that a 'last, non-direct click' attribution approach would tend to favour search over display advertising.

96.   Where an advertiser is only using one or two marketing channels or the route to purchase is reasonably direct, then a 'last touch' approach may well be sufficient to capture the key drivers of conversions. However, where there are multiple channels involved in a campaign – paid search, organic search, paid social (eg display advertising on Facebook), messaging on Twitter, e-mail targeting etc – then 'last touch' or 'first touch' attribution models would not capture the interaction of the different marketing channels.[24]

Multi-touch attribution

97.   Where an advertiser is making use of a number of marketing channels, they may use a 'multi-touch' or 'mixed-touch' attribution approach. This broad category of models recognises that a consumer may have interactions with a number of different marketing channels before converting. The difference between the different multi-touch attribution models comes in what value or weight is given to the different 'touchpoints' along the consumer journey.

98.   For instance, a 'linear attribution' approach would give equal weight to each interaction that led to the customer conversion. A 'time decay' approach would give more conversion weight to the interactions that happened closer to the conversion event. Another approach is a 'position based' attribution model, for example, 40% of the conversion is allocated to the first touch, 40% is allocated to the last touch and the remaining 20% is allocated across all other intermediate interactions that led to the conversion.

Data-driven attribution

99.   Advertisers do, of course, have the option to build their own attribution models and apply their own weights to the different touchpoints according to what they consider to be most relevant to their particular business strategy.

100.   The application of the different weights could be based on a combination of their own proprietary analysis (eg as a result of experimentation from switching off and on certain channels) or their own judgement. However, these more complex data-based attribution models are more difficult to construct.

---

[24] In addition to using multiple channels simultaneously, there can also be difficulties in attribution from a brand running ads on multiple websites concurrently or running multiple social media ad campaigns concurrently on the same site.

Attribution in practice

101. The ability to track a user across devices or sessions and to record their interactions with different adverts is an important part of measuring the effectiveness of digital advertising. It is also considered by advertisers to be an important component in being able to 'optimise' campaign performance in real time.[25] For instance, feedback from the attribution process about which adverts, or which marketing channels are having the most impact in terms of generating clicks to websites can help advertisers to re-allocate the campaign budget towards the most effective marketing channels.

102. Exposure data (also referred to as ad interaction data or campaign data) is data about all the instances where an advertiser's ads have been shown to users. From an advertiser's perspective, exposure data should ideally contain granular event data for each user, such as a user ID, datetime of the event, what creative was shown, as well as measurement issues like whether and how the user interacted with the ad (eg how long did they watch the video for, was their volume on, did they hover over or click on the ad, etc). These data are collected using the technologies described above, such as JavaScript tags and SDKs.

103. User-level exposure data is useful for measuring the distinct number of people in the audience of, and applying frequency capping for, an ad campaign across multiple publishers.

104. However, it is challenging to accurately attribute users across different platforms. As well as a user having multiple IDs with different firms, the process of attribution will be further complicated by cookie churn – cookies being deleted at periodic intervals – and the fact that a single user may be making use of a number of different devices, each of which would have its own user ID.

105. 'Walled garden' platforms are able to combine the ability to accurately monitor conversions with the ability to track users across different devices and sessions and so attribute consumers' actions more accurately. For instance, Google has a specific advantage in that it is able to track users across the 53 customer facing services which it owns. Its access to mobile data from Android also gives it some ability to the actions of consumers offline (eg to identify store visits).

---

[25] Given that attribution does not measure causality, the term 'optimise' appears to be used in a loose sense in relation to achieving certain intermediate campaign metrics.

106. In the case of Facebook, its single-user login feature[26] gives it a significant and specific measurement advantage over more standard cookie-based approaches in that it is able to associate all exposures and conversions across devices and sessions with a specific user as opposed to a browser on a laptop which could be shared between different people in a family.[27] Facebook also receives data from consumers activities across a wide range of other websites and apps when those services are using Facebook Business Tools.

107. In our interim report, we noted Google's decision to prevent the DoubleClick user IDs being accessed by advertisers and agencies. In 2018, Google restricted access to its User IDs (the DoubleClick ID) by removing it from its Campaign Manager and DSPs log files and curtailed the availability user-level exposure data from ad campaigns. This meant that ad buyers could no longer extract data from the DoubleClick Campaign Manager for reporting on ad performance and ad attribution.

108. Google indicated that the DoubleClick ID could be tied to sensitive information like user search histories and could violate the strict data privacy requirements of GDPR.

109. Stakeholders on the buyer side suggested that stripping out the DoubleClick ID removed visibility about user activity within the DoubleClick ecosystem and made it almost impossible to compare ad performance between ads purchased through the Google ad tech stack and ads purchased through other intermediaries. It was also suggested that the change made independent ad attribution much more difficult.

Single transaction ID

110. In our interim report we also proposed a requirement to comply with a common transaction ID that was applicable to all intermediaries as a way of facilitating third-party attribution.

Responses to our interim report

111. Broadly speaking, there was support for the principle of a common industry standard to facilitate cross-platform measurement but there were a number of different interpretations as to how it might be applied. For the most part,

---

[26] Facebook requires users to log-in to Facebook each time they access the service on any device and browser.
[27] At the same time, we note that Facebook would also be able to collect a great deal of location data from various sources and Google's persistent logins do provide an advantage over smaller competitors.

responses argued that what was needed was a common user ID rather than a common transaction ID.

112. As indicated above, in its response to our interim report consultation, Google told us that its approach to attribution was driven by its obligations under the GDPR. It cautioned that imposing consistent transaction IDs would raise potential privacy concerns: it might enable advertisers to join Google's secure bid data with other information in a way that would allow individual users to be identified.

113. For its part, Facebook argued that it was enabling advertisers to move away from 'final click' metrics towards multi-touch attribution systems in order to create better advertising campaigns and assess effectiveness of advertising campaigns more accurately.

114. Further information requests to and discussions with media agencies have indicated that the removal of the DoubleClick user ID did have the effect of preventing attempts to carry out customised attribution modelling analysis independently ie outside of Google products such as Google Ads Data Hub or Google Analytics360.

Data 'clean rooms'

115. Borrowing from the terminology used in sterile manufacturing environments, both Facebook and Google offer data 'clean room' services which enable advertisers to access granular campaign data under controlled conditions. In these settings it is possible for advertisers to combine their own first-party datasets with Google or Facebook datasets to generate reports on campaign performance and to carry out attribution analyses.[28]

116. We recognise that – over time – these facilities could offer a way forward. However, feedback from media buying agencies indicates that this is still a nascent area. At present the take-up of such services amongst their clients appears to be very limited and there are concerns that it is not possible to carry out analysis at the user level. For instance, we understand that the data available through Google Ads Data Hub ('ADH') still involves some level of user aggregation by Google.

---

[28] We also note that Amazon is reported as having begun to trial the use of its own version of a data clean room. See: Adexchanger: Amazon is Testing A Clean Room Service: Giving Advertisers Access to New Data Sets.

117. Furthermore, these products are platform specific and advertisers cannot carry out cross-platform analysis. The lack of the ability to extract data from these environments is also considered to be a constraint on their usage.[29]

Our assessment on attribution

118. Attribution is used by advertisers to make decisions about the design and implementation of advertising campaigns. It is also used to inform decisions about how to allocate advertising expenditure across different digital marketing channels.

119. Attribution for digital advertising is challenging and not straight-forward, particularly given the range of marketing channels and points of interaction between consumers and marketing messages across the course of an advertising campaign.

120. Google and Facebook have an advantage in relation to attribution because they can offer robust and accurate attribution for activities carried out on their own 'walled garden' platforms. Also, their tracking solutions are widely distributed across many other websites.

121. As set out in Appendix F, rival platforms such as Microsoft and Amazon have access to some detailed high-quality data about consumers and to other types of data, but this does not extend widely to the rest of the internet. This more limited access to data in terms of quantity and/or quality of campaign data on Google and Facebook's properties constitutes a barrier to entry and expansion, albeit not insurmountable, for smaller rivals in the provision of personalised advertising. In particular, lack of access inhibits independent providers of attribution services, and it could make it more difficult for advertisers to compare the relative performance of ads on Google and Facebook against ads on other websites and apps. Therefore, restrictions on third-party access to granular campaign data on Google and Facebook's properties may give Google and Facebook a competitive advantage.

122. Where those platforms (i) make it more difficult for third parties to implement their own attribution solutions (eg by removing or preventing access to the underlying user data necessary to carry out independent attribution analysis) and (ii) increase the reliance of their customers on the analytical products and services that they offer, then that can give rise to concerns about the exploitation of market power. Without the ability to carry out independent

---

[29] As indicated above, we are only aware of platform-specific data clean rooms. We are not aware of any third-party providers seeking to offer an independent option in terms of facilitating the analysis of campaign data from both advertiser and platforms.

attribution, there is a risk that advertisers are tied into to using Google and Facebook's own evaluation tools. As a result, they could end up over-paying for advertising purchased from those platforms and so mis-allocating their advertising expenditure relative to other sources of supply.

123. Both Google and Facebook restrict access to granular exposure data within their ecosystem. Google Campaign Manager (its advertiser ad server) has no API to export exposure data and generally prohibits exposure data from leaving Google. Using Google's ADH is the only way to access all of Google's post-campaign data and measurement analytics services in a single environment, and ADH does not permit customers to export the information to any other measurement or ad tech partner they may prefer.

124. These actions by Google have placed pressure on advertisers to choose between:

    (a) using Google for all of its ad tech and analytics services (Google ADH and Google Analytics), relying on aggregated reports from Google, and trusting that they are accurate, objective and complete; or

    (b) using independent ad tech (including a non-Google advertiser ad server) and analytics services but losing the ability to independently measure and verify campaign effectiveness across all of its impressions on multiple publishers in a unified way, assuming that advertisers would need to spend at least part of their budgets on Google and Facebook's inventory or face significant switching costs (discussed in Appendix M in the section on Competition between advertiser ad servers) and loss of supply.[30]

125. Some providers also told us that the removal of user-level campaign data inhibits advertisers' ability to compare – to some extent – the relative performance of Google's DSP and rival DSPs, particularly in terms of attributing conversions to multiple ads that the user may have been exposed to across different properties (multi-touch attribution, as opposed to last-touch attribution).

126. We recognise that the removal of access to user level data for attribution purposes has been linked to privacy concerns and, in particular, the introduction of the GDPR.

127. Responses to our interim report indicate that there is general industry agreement that a standard way of identifying a user across platforms would improve the accuracy and robustness of measuring the effectiveness of digital

---

[30] A possible third option would be for advertisers to forgo attribution altogether.

advertising. At the same time, there does not seem to be industry agreement as to how that could work in practice.

Measuring the effectiveness of digital advertising

Introduction

128. In theory, the availability of significant volumes of user-level data on the delivery of personalised digital advertising, combined with the ability to track follow-on purchasing behaviour, should mean that it is possible to measure the effectiveness of advertising much more precisely. Indeed, Goldfarb and Tucker (2011)[31] attribute the success of online advertising to its superiority over other advertising media in terms of its measurability and its targetability. This should in turn lead to a more efficient allocation of marketing budgets between different platforms and different forms of advertising.

129. However, the fact that advertising typically explains only a small proportion of the variance in purchasing,[32] combined with the fact that personalised digital advertising is heavily targeted, can give rise to significant identification and endogeneity problems. As a result, standard industry observational methodologies[33] (such as regression analysis) can easily generate biased estimates of the effectiveness of digital advertising which makes measuring the true incremental effect of advertising still very challenging.

Concerns about measuring effectiveness

130. A common problem with observational measurement approaches is that they implicitly assume that clicking on an ad and making a purchase is a caused by the consumer being exposed to that advert. In the case of personalised digital advertising this is a questionable assumption because of the degree of targeting involved. That is, the advertising has been targeted at customers who are considered likely to purchase the advertised product in the first place.

131. This is an example of endogeneity in that the digital advertising could – wrongly – be regarded as having caused the consumer to purchase the product when in fact the advertising was targeted at the consumer because they were likely to purchase the product. The consumer may have been intending to make a purchase anyway and so the advert had no incremental

---

[31] Goldfarb, A. and Tucker, C.E. (2011) 'Privacy Regulation and Online Advertising,' Management Science, Vol. 57, No. 1, pp. 57–71.

[32] Individual sales outcomes are volatile relative to ad spend per customer for many advertisers. For instance, car sales can be either £0 or many tens of £000s.

[33] In this context 'observational studies' refer to observing consumer behaviour in the presence of the advertising over a period of time ie without a control group.

impact on the purchasing decision. This 'selection bias' is likely to be a particular issue in relation to re-targeting in digital advertising where ads are served to consumers that have already visited a firm's website or is already on the firm's database of previous customers.

132.   Several studies (Lewis et al (2015), Gordon et al (2018)) have identified a number of ways in which standard ways of measuring the effectiveness of advertising could suffer from problems of 'endogeneity', leading to biased results. These include:

(a)   User-induced endogeneity (or 'activity bias'): to be exposed to digital advertising, a consumer has to be browsing online in the first place. As a result, more active users tend to see more ads. In addition, more active users are likely to carry out more activities online in general eg purchasing products, clicking on links and registering for information. As a result, exposure to digital advertising will be highly correlated with many online activities.

(b)   Targeting-induced endogeneity: modern advertising systems optimise who are shown ads. As an advertising campaign progresses, the systems learn which types of users are most likely to meet the campaign objectives (eg click through) and start to favour showing ads to those users that are likely to have a high click-through rate.

(c)   Competition-induced endogeneity: winning an auction for a particular impression means that the advertiser outbid all other advertisers bidding for the same impression. That means that the advertiser's ad is more likely to be shown to users that the advertiser values highly eg because of a higher probability of converting.

133.   Unless these issues are addressed, there is a risk that estimates of the effectiveness of digital advertising campaigns as measured by standard observational methods (such as regression analysis, propensity score matching) would suffer from significant problems of bias. As a result, advertisers would not be able to evaluate the effectiveness of their advertising on a robust basis. This would have the potential to lead to a significant misallocation of expenditure between different advertising media and advertisers paying too much for advertising as a result of biased measurement.

134.   There is a general recognition across the industry that Randomised Control Trials ('RCTs') represent the 'gold standard' in terms of addressing these

concerns.[34] In its simplest form, an RCT involves randomly dividing a target audience into two groups: a treatment group that has the opportunity to see the advertising and a control group that doesn't. This is often termed an 'Intent to Treat' approach.[35] The effectiveness of the advertising can then be assessed by comparing the impact on the two groups.

135.   However, in practice RCTs can be difficult to implement. There can be technical issues in relation to 'cookie churn' and multiple device usage which can make it hard to maintain the integrity of an RCT in an online advertising setting (Gordon et al, 2019). Given that advertising may only explain a relatively small proportion of sales, the also means that even cleanly designed experiments need to be done at scale to be able to measure economically meaningful effects of advertising, let along generate precise confidence intervals.

136.   There has also been a perception that they are costly to set up (both in terms of the actual costs and the opportunity costs of not being able to advertise to the control group) and they can also take time to produce the analysis.

The scale of the problem

137.   If there was not much difference between observational methods and RCTs in terms of measuring the effectiveness of advertising this might not be too much of a problem. However, this is not the case.

138.   At one extreme, Blake et al (2014) reported the results of a series of large-scale field experiments at eBay to measure the causal effect of paid-for search advertising. A key finding of this paper was that returns from paid search were only a fraction of conventional, non-experimental estimates and that – for eBay – the effect of paid-for search marketing was at best limited. Indeed, in the extreme they found that brand-keyword advertising (eg 'eBay shoes) had no measurable short-term benefits at all and in some cases the returns from search advertising could even be negative.[36]

---

[34] For instance, on the industry side, the IPA's The Expert Guide to Measuring Not Counting: How to Evaluate Social Media for Marketing Communications (2015) states: 'Establishing and measuring causality requires careful construction of counter-factual scenarios, ie what would have happened without the marketing intervention. The gold standard in causality is randomised controlled trials.' (Para 3.7). Similarly, in the academic research Imbens and Rubin (2015) state that 'running a randomized experiment is still regarded as the "gold standard" for estimating causal effects since randomization leads to groups which are probabilistically equivalent on all potential confounding factors'.

[35] In the context of digital advertising, implementing an Intent to Treat approach requires a significant number of participants to generate robust results because of the need to ensure that the composition of the control and treatment groups is comparable.

[36] It should be noted that a lesser reported finding of the Blake et al (2014) study was that search engine marketing could still have a positive effect on the purchasing behaviour for new and infrequent users. That is, the

139. The results of the Blake et al (2014) study were reported extensively in trade press as well as other more general publications. Furthermore, it was reported that eBay had subsequently stopped advertising on branded keywords and had saved in the order of $50m yearly in marketing expenditure (Rao and Simonov, 2016).

140. To explore this issue in more depth and to address the challenge that the results of Blake et al (2014) were due to a very specific set of circumstances particular to eBay at the time, more recent studies have sought to quantify the potential scale of the measurement problem across a broader range of industries.

141. Gordon et al (2018) analysed the results of 15 large scale[37] advertising campaigns conducted on Facebook using RCTs over the period January – September 2015.[38] The 15 RCTs involved a range of sectors – retail, financial services, telecoms and e-commerce – and were used to establish a baseline for the actual effectiveness of the advertising campaigns.

142. Comparing the results of these RCTs with a number of standard industry observational methods, the study found that, in general terms, observational methods were generally inefficient in accurately measuring the causal effect of digital advertising. The extent of the measurement bias could be significant. For half of the 15 campaigns, observational methods overstated the percentage increase in purchase outcomes by more than a factor of three. That is, there was a difference of over 300% between the point estimate of the incremental uplift in sales as measured by observational methods and the actual results derived from RCTs. However, given the relatively small number of RCTs involved in this study, the authors reported that they were not able to identify the campaign characteristics that were associated with strong biases.

143. The study did indicate that the extent of the measurement bias was less pronounced for user registration and website visits. That is, standard observational methods were capable of generating results which were close to those produced by RCTs where the objective was registration or landing page visits. The explanation for this is that these outcomes are easier to trigger via an advertisement compared to a purchase.

---

results did support the idea that paid-for search advertising could still be important for smaller companies without well-known brands.

[37] The number of users involved in these RCTs varied from 2 to 140 million.

[38] The authors noted that the sample of campaigns selected were not chosen to be representative of all the advertising campaigns run across Facebook. However, it was chosen to cover a range of sectors: retail, financial services telecoms, tech and e-commerce.

'Ghost ads' RCT methodology

144. As indicated above, in order to address concerns about deficiencies in standard observational approaches, there was a recognition that RCTs offered a way of producing accurate and robust estimates of the incremental impact of digital advertising, but – as noted above – they were regarded as being difficult and costly to implement.

145. To address some of these issues, Johnson et al (2017) set out what has been termed the 'Ghost Ads' methodology for RCTs. As with an Intent-to-Treat RCT approach, the basic idea of the Ghost Ads approach is that users are randomly assigned to the treatment group or the control group and the control group is not exposed to the advert being studied. However, where users in the treatment group are exposed to the advert being studied – the 'focal advert' – users in the control group are exposed instead to the advert that would have won the auction to determine which advert should be shown *but for* the presence of the focal advert.[39] In the latter case the focal advert becomes the 'ghost ad' in that it was not shown to the control group.

146. The result of this approach is that it is able to take account not only of the direct uplift from the advert but also the lost sales that have been avoided if another product had been advertised instead. Johnson at el (2017) explicitly discuss the fact that advertising can play a 'defensive' role that blocks competitors from stealing away consumers as well as an 'offensive' role in terms of pulling in consumers.

147. A further attraction of the Ghost Ads methodology is that it does not rely on such large-scale sample sizes in the way that an Intent-to-Treat approach does and so should be cheaper to implement.

148. The study by Johnson et al (2017) was able to demonstrate – using the Ghost Ads methodology – that re-targeting digital advertising had a positive, causal effect on visits to websites and purchasing. This was an important result in that re-targeting campaigns are ones involving consumers who have already been exposed to the product being advertised. This means that they are considered even more likely to purchase and so the endogeneity problem could have been even more acute.

---

[39] For instance, the methodology simulates the ad platform's allocation mechanism eg a real-time auction bidding mechanism.

149.   Google itself has high-lighted the potential of the 'Ghost Ads' approach and described it as 'an ideal solution' although it has also acknowledged that building this 'ideal' is not easy.[40]

Responses to our interim report

150.   In our interim report we indicated that we were aware of the academic research that had raised concerns about the effectiveness of digital advertising.[41] We raised the concern that platforms with market power could take steps to reduce the degree of transparency in digital advertising markets, or refrain from taking steps to make them more transparent, forcing advertisers to rely on information and metrics provided by those platforms.

151.   Stakeholders did not choose to respond directly to this issue in our interim report. We followed up this issue through discussions and information requests to media agencies and further information requests to Google and Facebook.

Platforms

152.   Both Google and Facebook do offer advertisers the ability to measure the effectiveness of advertising using different forms of RCTs.

153.   In the case of Google, it offers a Conversion Lift test which is based on the 'Ghost Ads' methodology. However, this form of testing is limited to certain advertising formats, including: (i) display campaigns bought via auction; (ii) YouTube ads bought via auction (including TrueView for Action and Shopping); and (iii) bumper Ads. The Ghost Ads methodology is not supported for Google Search.

154.   Information provided by Google indicates that take-up of the Conversion Lift tool has been very modest in terms of both the number of advertisers and the amount of advertising revenue involved (ie less than £1 million in any one year).[42]

---

[40] Google 'A revolution in measuring ad effectiveness' (May 2015)
[41] CMA Digital Platforms and Online Advertising: Market Study Interim Report. Paragraphs 5.42-5.43.
[42] As set out in the main report, it is estimated that [200-250,000] advertisers used Google in 2019. However, the Conversion Lift tool would only be relevant to the larger advertisers for the larger advertising campaigns and, as indicated, this tool is not available for Google Search. Even so, that still implies [over £500m] was spent on display advertising, meaning that the amount being subject to testing appears to be very low.

Table O.1: Data on UK advertisers using Google's Conversion Lift tool

Year	No. of UK advertisers using Google's Conversion Lift	UK Advertising Spend [(1)(2)] tested using Conversion Lift I (in £)
2017	[0-10]	[£100-200,000]
2018	[30-40]	[£600-700,000]
2019	[10-20]	[£400-500,000]

Note: (1) Figures rounded to the nearest £1,000. (2) We have converted annual ad spend expressed in US$ from US$ to GBP using the Bank of England reported US$ into GBP annual average spot exchange rate (XUAAUSS)
Source: Submitted to the CMA by Google in response to a request for information.

155.   Academic research submitted by Google indicates that it has been researching and developing geo-targeting approaches as a way of being able to offer clients an experimental methodology that offers robust results and that can be implemented reasonably easily. In a geo-targeting experiment, a geographic area becomes the experimental unit rather than individual users and advertising spend is increased in a region or regions while held constant or reduced in other regions. Vaver and Koehler (2011) describe the application of geo experiments arguing that they are conceptually simple, have a systematic and effective design process, and provide results that are easy to interpret[43] They are also privacy friendly in that the outcomes of interest are aggregated across each geographic region in the experiment. No individual user-level information is required for the 'pure' geo experiments (Kerman et al, 2017).[44]

156.   As a result of this research, Google is reported as having expertise in using geographical experiments to measure the causal impact of increased advertising (Owen and Launay, 2016)[45] and that geo-experiments are now the standard tool for the causal measurement of online advertising at Google (Chen and Au, 2019).[46]

157.   We asked Google for data on the extent to which UK advertisers were making use of geo-experiments. That data is set out below. It should be noted that data for 2018 was only available from October 2018 onwards.

---

[43] Vaver, J., and Koehler, J. (2011). Measuring Ad Effectiveness Using Geo Experiments.
[44] Kerman, J., Wang, P., and Vaver, J. (2017). Estimating Ad Effectiveness using Geo Experiments in a Time-Based Regression Framework.
[45] Owen, A. B. and Launay, T. (2016) Multibrand geographic experiments.
[46] Chen, A. and Au, T.C. (2019) Robust Causal Inference for Incremental Returns on Ad Spend with Randomized Paired Geo Experiments.

Table O.2: Data on No. of experiments set up using Google's geo-experiments tool and associated ad spend by UK advertisers

Year	No. of Experiments	Sum of Ad spend tested (£) [1]
2018 [2]	[[0-10]	[£600-700,000]
2019	[30-40]	[£3.2-3.3m]

Notes:
(1) Figures rounded to nearest £1,000. Ad spend refers to absolute value of testing including both increases and decreases in ad.spend.
(2) Data only available from Oct. 2018 onwards.
Source: Submitted to the CMA by Google in response to a request for information. .

158. Google has explained that the data on geo-experiments set out in the table only refers to experiments set up using Google's geo-experiments tool and it does not know whether those experiments were actually implemented. Furthermore, although Google can support advertisers in designing, executing, analysing and controlling geo-experiments, advertisers can also conduct their own geo experiments completely independently (by geo-targeting their campaigns or using third parties). In addition, Google's design and analysis methodologies are available open source.

159. In the case of Facebook, its Conversion Lift tool is based on an 'Intent to Treat' approach.[47] Evidence from Facebook indicates that – compared to Google - not only have significantly more advertisers used this tool but also the amount of advertising revenue subject to this testing has also been significantly greater although still only a relatively modest proportion of overall ad spend with Facebook.[48]

Table O.3: Data on UK using Conversion Lift Testing on Facebook

Year	No. of UK-based advertisers using Conversion Lift testing	UK advertiser spend exposed to Conversion Lift testing (£) [1] [2]
2015	[30-40]	-
2016	[70-80]	-
2017	[100-150]	[£10-20m]
2018	[900-1,000]	[£20-30m]
2019	[700-800]	[£40-50m]

Note: (1) Figures rounded to the nearest £1 million. (2) We have converted ad spend expressed in US$ from US$ to GBP using the Bank of England reported US$ into GBP annual average spot exchange rate (XUAAUSS).
Source: Submitted to the CMA by Facebook in response to a request for information.

160. Gordon et al (2018) has described Facebook's ability to track users via a 'single-user login' across devices and sessions as representing a significant measurement advantage over more common 'cookie-based' approaches. The single-user login not only helps to ensure the integrity of the random assignment process, preventing users from being inadvertently shown an ad, but also it means that Facebook can associate all exposures and conversions

---

[47] Facebook argues that an Intent to Treat methodology currently offers a more optimal trade-off between bias and precision than a Ghost Ads methodology for the Facebook platform and that it would be challenging to implement this methodology. It suggests that the Ghost Ads methodology may be more suited to Google's Search platform.

[48] As set out in Chapter 2, it is estimated that the total spend on online display advertising in the UK in 2019 was £5.4bn of which we estimate more than half went to Facebook.

across devices and sessions with a particular user. This cross-device tracking can be critical because users will frequently be exposed to advertising on a mobile device but then might subsequently convert on a tablet or desktop computer.

161.   The comparison of the use of Conversion Lift testing between Google and Facebook suggests that advertisers are prepared to make use of Conversion Lift testing where it is available and where the integrity of the RCT can be maintained.

162.   We note that although Google has indicated that it has been developing an expertise in geo-experiments and that geo experiments are now a standard tool for the causal measurement of online advertising at Google, the use of geo-experiments by UK advertisers still appears to be very low, especially given the sums of money spent with Google. We note the possibility that advertisers could be carrying out geo-experiments independently of Google's specific geo-experiment tool. However, given its claim that geo-experiments are a standard tool for the causal measurement of advertising, we would have expected Google to be able to provide more evidence to support that claim.[49]

Advertisers and Media Agencies

163.   Information provided by advertisers, discussions with media agencies and subsequent RFIs have established that advertisers and media agencies are aware of the importance of focusing on the incremental impact of digital advertising. Both Facebook and Google also provide marketing materials which set out the importance of establishing the incremental impact of digital advertising.[50]

164.   Agencies have also indicated that there that there is a general awareness of the potential measurement issues associated with standard observational methods and that advertisers do make use of RCTs as a way of measuring the effectiveness of advertising campaigns. The use of experimental approaches includes geo-experiments as well as the use of Conversion Lift tests. Some agencies reported that up to three-quarters of their clients had made use of RCTs of some form.

165.   However, those discussions also indicate that RCTs are not necessarily deployed on a frequent or regular basis. The reasons for this include cost –

---

[49] We note that Rao and Simonov (2016) were able to identify experimentation in their data on levels of advertising expenditure by looking for sharp changes in the levels of advertising which could reflect active experimentation, or 'natural experiments' due to budget exhaustion, churn in marketing personnel and so forth.
[50] For instance, Facebook's Learn What Drives Results for Your Business using Incrementality Measurement, January 2018. And Google's A revolution in measuring ad effectiveness (May 2015).

advertisers wanting to maximise the amount of money going on the advertising itself – and scale. Agencies also suggested that it can take some time to extract the insights from any RCTs and that advertisers may need some time to apply those insights across other campaigns they are planning before engaging in re-testing. For instance, if an RCT established that a particular type of campaign had been effective, then it was unlikely that the advertiser would also run RCTs on similar campaigns, at least for some time.

166. Furthermore, given the product or sector, agencies also suggested that RCTs may not always be suitable as a means of evaluating the effectiveness of a campaign. For example, in cases where the majority of purchases are offline or where the 'conversion journey' can take a long period of time, the argument was that the validity of RCTs could be compromised (because conversion tracking could be degraded). It was also generally the case that RCTs tended to be limited to larger advertisers for larger scale campaigns and where they had the in-house resources to properly analyse and interpret the results.

167. One agency did suggest that some clients were wary about the independence of RCTs on the Facebook platform, perhaps reflecting some of the trust issues referred to earlier. However, the same agency also described the Facebook platform as being the most advanced for this type of testing.

168. There was also the suggestion that it can be more difficult to carry out RCTs outside of walled garden platforms.

169. Overall, it appears to be the case that although advertisers are aware of RCTs as the 'gold standard', in practice, they tend to view them as one of a number of different ways of evaluating the effectiveness of a campaign. In some instances, advertisers may make use of other, simpler experimental approaches (such as A / B testing) to evaluate effectiveness. Advertisers do also continue to make use of observational methods – such as econometrics – with appropriate caveats and other metrics to assessing the performance of advertising campaign in the round eg click-through-rate, CPC, CPA etc. Advertisers appear to rely on a mix of metrics (including metrics that focus on the delivery of an advertising campaign) and the exercise of their judgement rather than to rely on any single metric.[51]

Our assessment on measuring effectiveness

170. The information available to us indicates that advertisers and agencies are aware of the need to focus on the incremental impact of the digital advertising

---

[51] This was sometimes expressed to the CMA as there being 'no single point of truth' in relation to measuring the effectiveness of digital advertising.

that they purchase and are aware that experimental approaches offer a more robust way to do this than observational approaches.

171.   Both Google and Facebook make available a range of provide experimental tools to advertisers and agencies if they wish to use them, although the Ghost Ads experimental approach is not supported for Google Search.

172.   The evidence indicates that that advertisers do make use of experimental approaches and in particular Conversion Lift analysis tools. However, this practice does not appear to be as widespread as might be expected, given the sums of money invested in digital advertising.

173.   There are some practical constraints on their deployment on a widespread basis. For instance, it may be difficult to implement RCTs across all advertising campaigns and RCTs may not be available across all advertising formats.

174.   Moreover, advertisers do not typically rely on just one measurement approach to evaluate campaign effectiveness. Instead they will make use of a range of different metrics depending on the nature of the advertising campaign.

175.   Overall, techniques have been developed to overcome key measurement issues and Google and Facebook do make available the tools to assess the effectiveness of advertising on their platforms. However, our analysis does point to the importance of advertisers and agencies being able to evaluate the effectiveness of digital advertising and to continue to engage with platforms on the tools and metrics that they need access to in order to be able to do this on an independent basis.

Annex – academic research on measurement issues

176.   The issue of the measurement of the effectiveness of digital advertising is one which has been subject to academic research since the late 2000s. However, perhaps the key paper which raised the profile of the issue of the effectiveness of digital advertising was that by Blake et al (2014). The results of this paper were covered extensively in trade press as well as other more general publications: Harvard Business Review, The Economist etc.

Blake, Nosko and Tadelis (2014)

177.   The paper reported the results of a series of large-scale field experiments at eBay to measure the causal effect of paid search advertising. The experiments first involved the effect of stopping advertising of 'eBay' on Yahoo and MSN but keeping advertising on Google. They then looked at stopping bidding on non-branded keywords in certain areas of the US for a period of time (60-days).

178.   A key result from this work was that the returns from paid search were only a fraction of conventional, non-experimental estimates and that for well-known brands – such as eBay – the effect of search engine marketing was found to be – at best – limited. Indeed, in the extreme they found that brand-keyword advertising (eg 'eBay shoes') had no measurable short-term benefits at all and in some cases the returns from search advertising could even be negative. Off the back of this research, it was subsequently reported (Rao and Simonov, 2016) that eBay had stopped advertising on branded keywords and saved in the order of $50 million p.a. in marketing expenditure.

179.   However, a lesser reported finding of the same study was that search engine marketing could still have a positive effect on the purchasing behaviour for new and infrequent users. Blake et al (2014) suggested that paid-for search advertising could still have an important role for play for smaller companies without well-known brands.

Subsequent academic research

180.   Off the back of Blake et al (2014), subsequent academic research explored two main themes. The first theme focused on establishing whether the results of the eBay experiments held true more widely and – linked to that – sought to understand how advertisers responded to information about the efficacy of advertising. The second them focused on ways of improving the efficiency of experimental methodologies to enable advertisers to measure advertising

effectiveness robustly and accurately. Both themes considered the underlying reasons behind the measurement issues.

Theme 1

181.   Lewis and Rao (2015) was a meta study of 25 large field experiments using advertising campaigns run on Yahoo! over the period 2007-2011. The campaigns involved major US retailers and financial companies. The starting point of this study was that the effectiveness of advertising on purchasing behaviour in general had been an open question for a number of years but that digital advertising had given the impression that firms would be able to measure the returns to advertising more accurately.

182.   The study established that the targeted nature of digital advertising created a significant inference challenge for standard observational approaches and that experimental methods (eg RCTs) represented a more robust way forward. However, to be informative – to be able to distinguish which campaigns had a positive impact on purchasing behaviour – they found that RCTs needed to be of such a large scale that it meant that experiments were costly and potentially infeasible to many firms. They reported that all of the experiments they looked at were underpowered in terms of being able to distinguish between a RoI of 0% and one of 10%.

183.   As a result, their main finding was that most advertisers could not know the effectiveness of their advertising spend. However, they did acknowledge that their results were not likely to apply to firms or products with low baseline sales volatility such as those which generated most of their sales from advertising or had a low level of awareness.

184.   They did make two other relevant comments. The first was to suggest that larger publishers could enjoy a strategic, scale advantage because of the size of RCTs needed. The second was to suggest that imprecise signals on returns to advertising could raise issues of strategic mis-reporting *within* firms. That is, a marketing department might not have an incentive to draw attention to concerns about the efficacy of the advertising they commission with other parts of their organisation (eg the finance department).

185.   Simonov et al (2016) carried out a series of large-scale experiments on Bing studying some 2,500 brands. They did report a positive and statistically significant impact of brand advertising but only of a relatively modest size: an RoI of 14% compared to the more substantial RoIs that had been generated by standard observational methods. They also reported that there was

significant heterogeneity in this result: the impact for well-known brands showed a smaller causal effect.

186. Rao and Simonov (2016) examined five years of bidding data on Bing to examine the reaction of firms to the widely disseminated results of the Blake et al (2014) paper about the lack of effectiveness of brand advertising. They considered both the propensity of firms to advertise on own brand queries and to conduct experiments to measure the effectiveness of their advertising.

187. They found that only just over 10% of firms that did not face competing ads on their branded keywords – as in the case of eBay – actually stopped brand search advertising. In contrast, most firms continued with own-brand advertising despite the issues flagged in the Blake et al (2014) paper. They also examined whether there were any sharp changes in the levels of advertising expenditure which could suggest that firms were actively experimenting to test the impact of their advertising and so could be considered already to be optimising. That is, they had evaluating their advertising expenditure, were getting positive returns (in contrast to eBay) and had actively chosen to continue to advertise. In fact, they did not find any measurable impact on the propensity to conduct experiments.

188. The authors concluded that there was a difficulty in shifting beliefs about causal inference and speculated that this might be down to risk aversion. That is, people wanted to avoid the risk of being punished for 'past behaviour' ie having spent money on advertising whose efficacy was now being challenged. Along the same lines, they flagged a potential principal-agent issue between the marketing and finance departments within a firm.

Theme 2

189. Lewis et al (2015) discusses some of the key factors which impact on the measurement of the effectiveness of digital advertising.

190. In theory, the prevalence of individual-level data on the delivery of digital advertising and the ability to monitor follow-on purchasing behaviour, should mean that it is possible to measure the effectiveness of advertising much more precisely. In turn that should lead to a more efficiency allocation of marketing budgets between different forms of advertising.

191.   However, the fact that advertising typically explains only a small proportion of the variance in purchasing[52] and the fact digital advertising is heavily targeted gives rise to significant identification and endogeneity problems. As a result, industry standard observational methodologies can easily generate biased estimates because of unobserved heterogeneity. Even large scale RCTs can suffer from problems in terms of addressing basic causal issues.

192.   The paper identifies activity bias as a form of selection bias and therefore a significant issue for measurement. To be exposed to digital advertising, one has to be browsing online so more active users tend to see more ads. In addition, more active users tend to more of everything online such as buying goods, clicking on links and registering for information. As a result, ad exposure is highly and non-causally correlated with many online activities.

193.   The paper also argues that there is the need to consider off-line as well as online sales. Focusing just on online measurements could induce a large negative bias in measuring returns to advertising.

Ghost ads: Johnson et al (2017a)

194.   The article outlines an RCT methodology which is able to address key methodological issues identified in Lewis et al (2015) as creating issues even for RCTs. The Ghost Ads methodology is one which is capable of reducing the cost; improving measurement precision compared to other RCT approaches (eg Public Service Announcements ('PSA')); and working with modern ad delivery platforms that optimise ad delivery in real time.

195.   The basic idea of the Ghost Ads approach is that the control group is exposed to the advert that would have won the auction *but for* the presence of the focal ad. It does not replace the focal advertisers' ad with a completely different type of ad (as happens with PSA experiments). The result of this approach is that it is able to take account not only of the direct uplift from the advert but also the lost sales that have been avoided if another product had been advertised instead. Johnson at el (2017a) explicitly discuss the fact that advertising can play a 'defensive' role that blocks competitors from stealing away consumers as well as an 'offensive' role in terms of pulling in consumers.

196.   The paper uses this approach to demonstrate that re-targeting digital advertising has a positive (causal) effect on visits to websites and purchasing.

---

[52] Individual sales outcomes are volatile relative to ad spend per customer for many advertisers. For instance, car sales can be either £0 or many tens of £000s. This means that even cleanly designed experiments require a very large number of subjects to be able to measure economically meaningful effects of advertising let along generate precise confidence intervals.

This is an important result in that re-targeting campaigns are ones involving consumers who have already been exposed to the advertising and are so considered even more likely to purchase. However, the estimates of the scale of advertising effectiveness is relatively modest compared to estimates produced by other methodologies.

197.   In a parallel working paper Johnson et al (2017b) examined the results 432 online display ad field experiments on the Google Display Network involving 431 advertisers across a range of industries. Again, this work pointed to digital advertising having a positive, causal effect on conversions but that this effect was not spectacular. Johnson et al (2017b) found both a positive in-campaign lift in terms of site visits and conversions. It also found a positive but modest carryover effect for four weeks after campaigns had ended.

Gordon et al (2018)

198.   Using the Ghost Ads methodology, Gordon et al (2018) took as its starting point the results of 15 large scale advertising campaigns conducted on Facebook as RCTs over the period January – September 2015.[53] These 15 RCTs established establish a baseline of the actual effectiveness of the advertising campaigns. In all but one of the studies, the effectiveness of the advertising campaigns was measured in terms of the impact on purchase outcomes. In a number of campaigns other measures of effectiveness – user registration and visits to a website – were also compared.

199.   The study benefitted from the fact that the sample size of the RCTs were large, involving between 2 and 140 million users. The data was also collected at the user level – rather than device or browser level – and Facebook's 'single-user' login feature meant that it is possible to track exposures and conversions across all of a user's devices.

200.   The study then used these 15 datasets to evaluate a number of different observational methods including matching and regression-based methods. The authors then compared the results of these observational approaches with the results from the RCTs.

201.   As a general point, the study does note that six of the RCT studies failed to produce a statistically significant sales uplift at the 5% significance level. That is, the RCT indicated that in those cases, it was not possible to show that the

---

[53] The authors noted that the sample of campaigns selected were not chosen to be representative of all the advertising campaigns run across Facebook. However, it was chosen to cover a range of sectors: retail, financial services telecoms, tech and e-commerce.

advertising campaign had had statistically significant impact on purchase behaviour.

202. The study found significant differences in the measures of ad effectiveness obtained from the RCTs compared to those from the observational methods. Generally, the observational methods overestimated the effectiveness relative to the RCT but in some cases they end up underestimating the effectiveness.

203. The extent of the measurement bias can be large. For half of the campaigns, the estimated percentage increase in purchase outcomes was off by a factor of 3 across all methods. That is, there was a difference of over 300% between the point estimate of the incremental uplift in sales as measured by the RCT and by the observational methods. However, with the relatively small number of studies, the authors reported that they were not able to identify the campaign characteristics that were associated with strong biases.

204. The study did, however, indicate that the extent of the measurement bias was less pronounced for user registration and website visits. That is, standard observational methods were capable of generating results which were close to those produced by RCTs where the objective was registration or landing page visits. The explanation for this is that these outcomes are easier to trigger via an advertisement compared to a purchase.

205. The authors also noted that, in some cases, with good data, observational methods were capable of yielding estimates which were similar to those generated by RCTs. However, none of the observational methods yielded consistent results relative to the RCT benchmark.

Bibliography

Blake, T., Nosko, C., and Tadelis, S. (2014) 'Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment.' NBER Working Paper 20171.

Gordon, B.R., Zettelmeyer, F., Bhargava, N., and Chapsky, D. (2018) 'A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook.'

Johnson, G.A., Lewis, R.A., and Nubbemeyer, E.I. (2017a) 'Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness.'

Johnson, G.A., Lewis, R.A., and Nubbemeyer, E.I. (2017b). 'The Online Display Ad Effectiveness Funnel & Carryover: lessons from 432 Field Experiments.' working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701578

Lewis, R., Rao, J.M., and Reiley, D.H. (2015) 'Measuring the Effects of Advertising. The Digital Frontier' in *Economic Analysis of the Digital Economy.*

Lewis, R.A., and Rao, J.M. (2015) 'The Unfavourable Economics of Measuring the Returns to Advertising.'

Rao, J.M., and Simonov, A. (2016). 'Firms' Reactions to Public Information on Business Practices: Case of Search Advertising.'

Simonov, A., Nosko. C., and Rao, J.M. (2015) 'Competition and Crowd-out for Brand Keywords in Sponsored Search'