

Quantitative testing of Google's Privacy Sandbox technologies – CMA guidance to third parties on testing

29 June 2023

Summary

Google intends to test quantitatively the effectiveness of its Privacy Sandbox technologies between Q4 2023 and Q2 2024 in order to inform the CMA's assessment of the tools. We are also seeking evidence from testing carried out by third party market participants.

In this note, we are seeking to advise ad techs, publishers, and advertisers on how they can test the Privacy Sandbox tools in a way that would contribute to our assessment of the Privacy Sandbox technologies.

It provides details of two preferred approaches to testing, the metrics we would like to capture, and information market participants can submit to the CMA so we can understand the results of their testing.

Introduction

1. Under its Privacy Sandbox proposals, Google plans to begin deprecating the use of third-party cookies ('**TPCs**') in Chrome in the third quarter of 2024 and replace many use cases they currently serve with alternative technologies, provided there are no competition concerns remaining. As part of legally binding commitments (the '**Commitments**') agreed with the Competition and Markets Authority (the '**CMA**') in February 2022, Google will quantitatively test the effectiveness of these alternative technologies against a set of criteria (the Development and Implementation Criteria), which include their impact on publishers, advertisers, and competition.¹
2. Alongside a range of other evidence, the results of Google's quantitative tests will inform the CMA's assessment of whether the Privacy Sandbox has been designed in a way that addresses our competition concerns.² However, the effectiveness of these technologies may vary across market participants, depending for example on the type of publishers/advertisers served or the

¹ The [Commitments](#), paragraph 8.

² [Commitments](#), paragraph 7. See also [Commitments Decision](#), paragraphs 3.29 to 3.83.

range of complementary signals and technologies used. For this reason, we would like to encourage other market participants to perform similar tests and provide the results to the CMA. Because they cover the market for ad tech services and represent publishers and advertisers, we view the tests we are proposing as particularly relevant for DSPs and SSPs. We encourage all of those who participate in the digital advertising market to consider whether and how they are able to engage in testing and contribute to our evidence base.

3. In November 2022, we published a note outlining and requesting feedback on proposals for experiments (or A/B tests) that third-party market participants might use to test the impact of the Privacy Sandbox on a range of business metrics (the '**experiments note**').³ Feedback on those proposals was broadly positive. However, market participants told us they required further clarity on the practicalities of testing and timelines to engage with them properly. They also reiterated that testing is costly and time-consuming.
4. Since the publication of the experiments note, we have continued to develop our thinking on how market participants can engage in testing in a way that minimises the additional burden it might place on their already constrained resources. In addition, Google has recently announced that, beginning in Q4 of this year, it will enable testing environments in Chrome that will support third-party testing.⁴ This note:
 - i. outlines the types of testing designs that are most likely to be informative for the purpose of assessing the effectiveness of the Privacy Sandbox APIs;
 - ii. describes how our proposed testing approaches fit with the support Google intends to provide for testing, including implications for timing; and
 - iii. outlines how market participants can report test results to the CMA in a meaningful way.

³ See [Quantitative testing of Google's Sandbox technologies](#), November 2022.

⁴ See the announcement here [The next stages of Privacy Sandbox: General availability and supporting scaled testing](#); and accompanying developer blog post here: [Preparing to ship the Privacy Sandbox relevance and measurement APIs - Chrome Developers](#).

Preferred approaches to testing the effectiveness of the Privacy Sandbox APIs

Overall approach and role of testing in our assessment

5. Our assessment of the Privacy Sandbox proposals will draw on a wide range of qualitative and quantitative evidence on whether Google has addressed our competition concerns.
6. Under the Commitments, Google committed to design, implement and evaluate the Privacy Sandbox proposals by taking into account certain factors (the 'Development and Implementation Criteria'), which will inform the answer to the question of whether or not the competition concerns have been addressed. Two of these Development and Implementation Criteria centre around the impact on competition and the impact on publishers.
7. These factors are primarily related to the *effectiveness* of the targeting and measurement APIs in replicating the functionality currently supported by TPCs.⁵ As described in our decision to accept the Commitments (the '**Commitments Decision**'), the assessment of these concerns could be supported by the results of experiments (or A/B tests).⁶ In the experiments note, we outlined two ways in which market participants could use experiments to measure how the Privacy Sandbox is impacting their competitive outcomes.
8. We do not intend to coordinate one industry-wide test of the Privacy Sandbox technologies, nor do we intend to propose use-case-specific tests that market participants should follow. We understand that many market participants are investing a lot of resources in preparing for the deprecation of TPCs in Chrome, and that conducting experiments that do not necessarily align with their own needs might be impractical. Instead, we encourage market participants to run tests along the lines of those outlined in our experiments note (and described below), while reflecting their own business models and constraints.
9. We also understand that market participants might be interested in testing alternatives to Google's Privacy Sandbox. While the provisions for quantitative testing in the Commitments relate specifically to the testing of Google's

⁵ See Commitments Decision, [Appendix 4](#), paragraph 13 for the difference between *functional* and *effectiveness* testing; and paragraph 12 for the three broad use cases Google has committed to consider in its quantitative testing.

⁶ Commitments Decision, [Appendix 4](#), paragraph 15(a). The experiments we describe in this note can also be referred to as A/B tests.

technologies, we would encourage market participants to carry out such tests and submit their results to the CMA – testing of alternatives to Google’s Privacy Sandbox can contribute to our wider assessment.

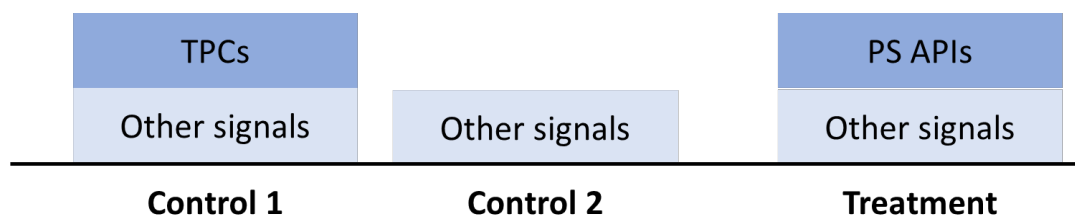
Proposed treatment and control groups

10. In general, the experiments we are proposing compare selected market outcomes for subsamples of impressions served using Privacy Sandbox technologies (**‘treatment impressions’**)⁷ against a subsample of ads served under counterfactual scenarios (**‘counterfactual impressions’** or **‘control impressions’**).
11. We propose market participants use the following counterfactual groups to make such comparisons:
 - i. **control group 1:** ads served using data related to TPCs and removing data related to new APIs before issuing the request for bids; and
 - ii. **control group 2:** ads served removing data related to both TPCs and the new APIs.
12. These control groups can be viewed as representing two extremes in a range of potential counterfactual scenarios: one in which current use of TPCs continues (control group 1); and another in which there are no replacement technologies after TPCs are deprecated (control group 2). It is not currently possible to define a precise counterfactual for the implementation of the Privacy Sandbox, given that such a counterfactual scenario could include some unknown future technology. We view the two control groups described in 11.i and 11.ii as representing bounds on market outcomes without the Privacy Sandbox. We do not intend to use comparisons between the treatment group and the two control groups in isolation in our assessment. Rather, we intend to use comparisons across the groups alongside wider quantitative and qualitative evidence to understand what impact the deprecation of TPCs and introduction of the Privacy Sandbox will have on competition.
13. We understand that market participants use a variety of other signals aside from TPCs to assign ads to ad requests, for example first party publisher data and contextual information. To the extent that these signals are not impacted by the proposed changes (the deprecation of TPCs and the introduction of the Privacy Sandbox APIs), these should be retained in both the control and

⁷ We understand that in the treatment group market participants might also use other replacement technologies available at the time of the experiments alongside the Privacy Sandbox APIs.

treatment groups. The diagram below summarises our proposed definition of the experiment groups.

Figure 1: an illustration of which signals should be used in the treatment and control groups



14. The mix and significance of these other signals, and the technologies used to incorporate them into bidding logics, may vary across market participants. In general, each party conducting experiments should aim to include the mix of signals that they are most likely to use in the future.

Proposed experiment designs

15. Making experimental comparisons between ads served with and without the Privacy Sandbox requires a mechanism for allocating ads to treatment and control groups. There are two possible approaches:
 - i. **Design 1:** participating market participants create their own treatment and control groups by randomly suppressing/retaining TPCs/Privacy Sandbox technologies in subsamples of ad requests.
 - ii. **Design 2:** the Chrome browser creates treatment and control groups by randomly suppressing/retaining TPCs/Privacy Sandbox technologies at the browser level. Participating SSPs can then bring impressions from the relevant browsers to auction, adding any other signals that they are using or are likely to use in the future.
16. The advantage of Design 1 is that it can be implemented relatively quickly by interested market participants without the need for any supporting role by the Chrome browser (for example, this is the approach taken by Google Ads in its recent experiment testing the impact of Topics on interest-based advertising⁸). The disadvantage of Design 1 is that it provides no guarantee that DSPs will bid in treatment auctions using just the Privacy Sandbox APIs (alongside other signals). For example, a DSP could bid on impressions placed in the treatment group using information from TPCs obtained from outside the

⁸ See the blog post here: [Results from Google Ads' interest-based advertising testing \(blog.google\)](https://blog.google/ads-privacy/Testing-IBA-with-Privacy-Preserving-Signals.pdf); and the accompanying whitepaper here: [ads-privacy/Testing IBA with Privacy Preserving Signals.pdf at master · google/ads-privacy · GitHub](https://ads-privacy/Testing-IBA-with-Privacy-Preserving-Signals.pdf).

auction. If this were the case, then experiments based on this design may not be indicative of the effectiveness of the Privacy Sandbox APIs.

17. By suppressing TPCs entirely at the browser level, Design 2 overcomes this issue, and for this reason we are likely to put more weight on results based on that design in our final assessment. However, results from experiments based on Design 1, if available earlier, might provide some useful information that could be incorporated into the design of the Privacy Sandbox APIs or the design of future experiments based on Design 2.
18. Participating SSPs and DSPs might also find it useful to work together on experiments, as appropriate. For example, participating SSPs might communicate the parameters (eg timing) of their experiments to DSPs in their stack, and encourage them to bid on impressions in the treatment group using signals related to the new APIs in a way that most closely approximates the way these signals would be used if they were available on a large scale (with TPCs deprecated). This could result in experimental auctions being more representative of potential post-TPC auctions, increasing their relevance for our assessment.

Key metrics and measuring impacts

19. We are interested in hearing about a range of outcomes resulting from the experiments. As we described in paragraph 7 above, experiments are well suited to testing the impact on advertisers, publishers, and competition, so we are most interested in results showing how the Privacy Sandbox might affect:
 - i. Revenues per impression
 - ii. Clicks and conversions per dollar
 - iii. Clicks per impression
 - iv. Web page latency
 - v. Total unique bid requests served by DSPs/SSPs (as a proxy demand for individual ad techs' services)⁹
 - vi. % of planned campaign budget spent
 - vii. Unique viewers

⁹ We are interested in both the number of unique ad requests received by and the amount of inventory bought/sold through ad tech services.

- viii. Average time spent or video completion rates
 - ix. Brand lift.
20. This list of metrics is not exhaustive, and market participants should test the impact of the Privacy Sandbox on metrics they consider important.¹⁰ We note that the metrics we have listed above are also only suggestions, and we understand that market participants might not have visibility into some of them.
 21. As regards conversions-based metrics, we understand that measuring the impact of the Privacy Sandbox on these will be affected by the complete deprecation of TPCs in the treatment arm in Design 2. Consequently, additional testing of the impact on conversions in Design 1 - where TPCs can be used for measurement - would be helpful.
 22. If it is possible, and market participants are able to run experiments over time, we propose collecting metrics over several months to understand whether and how the impact of the Privacy Sandbox is changing as the models used by market participants evolve.
 23. To estimate the impact of the Privacy Sandbox on the metrics in 19.i-ix to , market participants should compare average outcomes across treatment and control groups. For example, the impact on revenue per impression could be estimated as follows:

$$\Delta_{PS} = \overline{RPM}_{treatment} - \overline{RPM}_{control\ 1}$$
 24. Where $\overline{RPM}_{treatment}$ is the simple average of revenue per thousand impressions among impressions in the treatment group, and $\overline{RPM}_{control\ 1}$ is the same metric for control group 1. Market participants can also calculate impacts (ie Δ_{PS}) as percentage changes, or more complex statistics like Mantel-Haenszel (MH) ratio that Google calculated in its recent Topics IBA experiment.¹¹
 25. Because these effect estimates are based on sample averages, market participants should also calculate sampling variability in the form of standard errors (and/or confidence intervals). There are various methodologies

¹⁰ In the experiments note, we also included the following metrics: browser crashes per 1 million page loads and share of users closing ads. The former of these is a metric that Chrome is best placed to measure, and we understand that it will likely be infeasible for market participants to measure the latter by the time quantitative testing takes place. See paragraphs 31 to 33 for more details on the testing timetable.

¹¹ See section 4, footnote 2, and Appendix 3 of Google's paper on their recent experiment here: [ads-privacy/Testing IBA with Privacy Preserving Signals.pdf at master · google/ads-privacy · GitHub](#).

available to calculate standard errors, each of which are suitable in different setting. Market participants should select a methodology based on how they calculate their impacts (does the statistic have a well-know distribution) and the structure of their data (eg are there adjustments required for clustering). If unsure about the appropriate methodology to employ, market participants should feel free to consult with us for advice.

Information on the composition of treatment and control groups

26. In addition to the metrics listed above - which represent outcomes we would like to assess - we also request that market participants submit information on the characteristics of the impressions or users that make up their treatment and control groups.
27. This will allow us to assess whether the randomisation of impressions (in Design 1) or browsers (in Design 2) was effective in creating comparable subsamples (and whether these subsamples are reflective of the characteristics of the broader population of Chrome users), eg share of impressions in the in different countries/geographies.¹²
28. We encourage the submission of any relevant characteristics market participants can record that might allow us to assess whether treatment and control impressions are comparable.

Chrome-facilitated testing environments

Chrome's testing modes

29. To make it easier for market participants to experiment with the Privacy Sandbox tools using Designs 1 and 2, both independently and together where appropriate, Google recently announced plans to implement two testing modes in Chrome:¹³
 - i. **Mode A:** In this mode, TPCs will still be available but Chrome will assign a portion of traffic to treatment and control groups and provide ad techs with labels telling them to which group traffic belongs. We understand Google's current intention is to provide labels for up to 10% of Chrome

¹² We understand that the treatment and control groups might differ in ways that are not readily observable. For example, we are not be able to observe the characteristics of the consumers using the browsers. However, understanding the extent to which they are matched in terms of observable characteristics will allow us asses whether they are likely to be balanced

¹³ See the announcement here [The next stages of Privacy Sandbox: General availability and supporting scaled testing](#); and accompanying developer blog post here: [Preparing to ship the Privacy Sandbox relevance and measurement APIs - Chrome Developers](#).

browsers through a new request header and low-entropy client-hint. The state of TPCs will not be modified in Mode A, and traffic will only be labelled so market participants can experiment on the same treatment and control groups. Mode A is intended to facilitate experiments **along the lines of Design 1** (paragraph 15.i above) but with treatment and control groups defined by the Chrome browser rather than market participants themselves. Market participants can test independently or together with others where appropriate.

- ii. **Mode B:** Chrome will deprecate TPCs for 1% of traffic globally (to form the treatment group).¹⁴ A fraction of traffic will also have Privacy Sandbox relevance and measurement APIs disabled (to form control group 2).¹⁵ Because TPCs will be deprecated in Mode B, it is designed for ad techs to conduct experiments **along the lines of Design 2** (paragraph 15.ii above).
30. In both modes, market participants will be able to identify which ad requests are in treatment and control groups through Google's experimental labels. We understand that labelling traffic in this way 'unblinds' the experiment, in that participants will be able to see treatment/control allocations and might change their behaviour as a result. However, particularly during early testing in Mode A, it is important that market participants are able to deploy their models and resources effectively and efficiently on the correct traffic (ie using only Privacy Sandbox technologies in the treatment group and, eg TPCs for control group 1).¹⁶ Knowing to which group impressions belong also means market participants will be experimenting on the same impressions as Google, allowing for easier aggregation and comparison of results. We currently consider these benefits to outweigh the potential costs of signalling the experimental status of impressions.
31. Google has published more details on the two testing modes, the process for enrolling in testing, and how to work with the various Privacy Sandbox tools in its recent blog post on preparing to ship the relevance and measurement APIs.¹⁷ The CMA also welcomes feedback on Google's testing modes, how they align with the experimental designs we have proposed for third parties, and any challenges market participants foresee in using them.

¹⁴ Google are actively working on mitigations for any issues this small-scale deprecation might have on user experience.

¹⁵ Google is currently seeking feedback on what market participants consider and appropriate proportion of traffic for this treatment arm. See: [Preparing to ship the Privacy Sandbox relevance and measurement APIs - Chrome Developers](#).

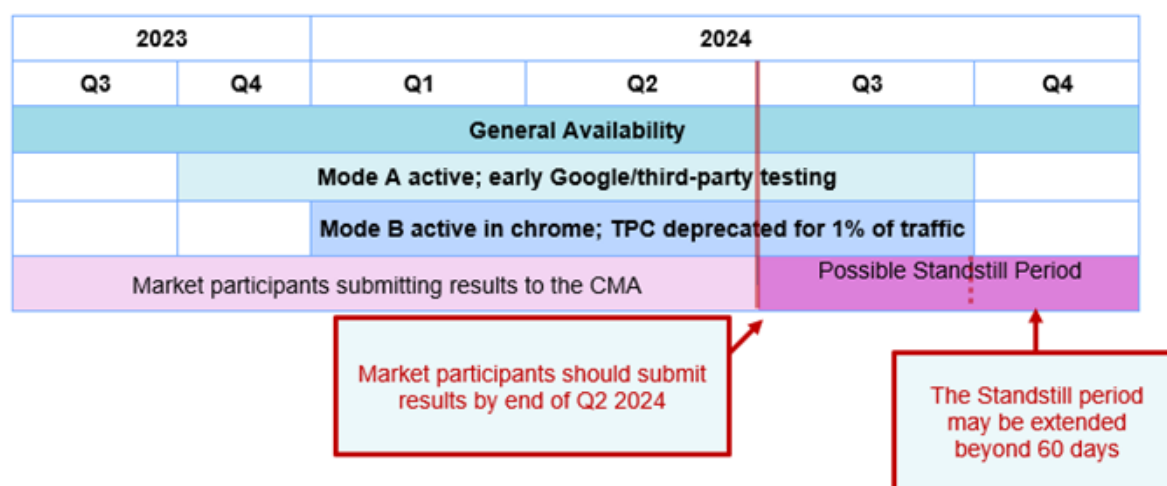
¹⁶ We understand that: (a) market participants each have their own technology/models; and (b) in the treatment group market participants might also use other replacement technologies available at the time of the experiments alongside the Privacy Sandbox APIs.

¹⁷ [Preparing to ship the Privacy Sandbox relevance and measurement APIs - Chrome Developers](#)

Testing timetable

32. The Stable Origin Trials ('OT') for Topics, FLEDGE (now Protected Audience), and Attribution Reporting APIs are currently live and are scheduled to run until the end of September 2023.¹⁸ From Q3 2023, Google will move the APIs to General Availability in preparation for its planned date of TPCs deprecation in Q3 of 2024 (which is subject to our competition concerns being addressed).¹⁹ Google intends to conduct its quantitative testing during the General Availability phase in order to, where possible, assess the effectiveness of the Privacy Sandbox against the Development and Implementation Criteria. The CMA encourages market participants to submit their results as early as possible, and at the latest **by the end of Q2 of 2024** to enable us to include them in our assessment during the Standstill Period.²⁰ Figure 2 below shows the testing timeline visually.

Figure 2: A visualisation of the CMA's proposed testing timeline and Google's Chrome facilitated testing modes



¹⁸ Google's Origin Trials will also be available until mid-September 2023. See [here: ps://developer.chrome.com/origintrials/#/register_trial/771241_436187197441](https://developer.chrome.com/origintrials/#/register_trial/771241_436187197441).

¹⁹ See 'The Privacy Sandbox Timeline for the Web'

²⁰ Under the [Commitments](#) (paragraph 19), Google will not remove TPCs before the expiry of a standstill period of no less than 60 days after Google notifies the CMA of its intention to implement their removal (the '**Standstill Period**'). We will perform our assessment of the Privacy Sandbox technologies during the Standstill Period to determine whether any competition concerns remain. The Standstill Period can be extended to a total of 120 days at the CMA's request.

33. Google also intends to align the Chrome-facilitated testing environments (described in 29.i and 29.ii above) with this testing timetable to allow Google and third-party market participants to experiment on the same ad impressions:
- i. **Mode A:** Google intends to make this mode available at the beginning of General Availability, from the start of Q4 2023,²¹ and keep it active until TPC deprecation in Q3 2024.
 - ii. **Mode B:** Google intends to make this mode available from Q1 2024 until TPC deprecation in Q3 2024.
34. The activation of Mode B will not end Mode A, and market participants can continue to test using Mode A for initial testing throughout General Availability and once Mode B becomes active. However, we would like both testing modes to be available for use as early as possible to allow testing across a range of market participants who might be at different stages of readiness for testing. We also understand that market participants might not be ready to run experiments as early as Q4 2023. However, we would like the testing modes to be available as early as possible for those who are ready to test. We also encourage market participants who are ready to run tests before either mode is ready, to do so.

What we are interested to hear about your tests

The information we need about your results

35. We are seeking to encourage a wide range of market participants to engage in quantitative testing. We understand that market participants differ in terms of their own technologies, how they intend to use the Privacy Sandbox, and, as a result, the types of tests that might be useful to them to perform. As part of our assessment during the Standstill Period, we would like to use results from these tests to assess the impact of the Privacy Sandbox on competitive outcomes in the market for digital advertising. To do this, we wish to understand a number of practical features of any tests conducted so we can:
- (a) more reliably compare test results across market participants; and
 - (b) understand what might be driving results or differences in results.

²¹ We understand that the launch of Mode A during Q4 of 2023 clashes with the Christmas holiday and new year period, which is also a particularly important retail period. However, we would like to facilitate testing as early as possible, and consider that tests using Mode A, particularly at an early stage, would be focussed less on measuring impacts on market outcomes. See paragraphs 16-18.

36. In 36.i-xiii below, we outline the questions market participants should answer when submitting test results to the CMA:
- i. **Was testing visible or invisible?** Was the test carried out on live inventory and did it result in: showing different ads or formats to end-users; or an impact on the prices paid for ad inventory?
 - ii. **Which inventory, formats, geography, and campaign types did your testing cover?** Did the test involve all current lines of customers, services, product offerings, and campaign types in your business that rely on TPCs? Or was it a partial test of customers, services, products, or campaign types?
 - iii. **How comprehensive were the changes you made for testing to your infrastructure?** Did the test use all, none, or some of the serving, bidding, billing, spam/fraud and reporting stacks you plan to use once TPCs are deprecated? Did you use pre and post-TPC stacks in combination, or use them on only a small fraction of traffic?
 - iv. **How comprehensive were any mitigations²² you used?** Did the test incorporate all, some, or no mitigations, and did it represent a comprehensive view of how you plan to incorporate different signals (like contextual of first-party publisher data) outside of the Privacy Sandbox when TPCs are deprecated? Eg, were any signals left out of the test. Could you provide a high-level list of the mitigations you used?
 - v. **How much did you rely on the mitigations you used?** Is it possible to say how much you relied on, for example, Topics relative to other mitigations?
 - vi. **How business-critical were the metrics you collected?** Do the metrics collected represent a comprehensive overview of the metrics you would use in your ordinary commercial activities, and which your customers would primarily rely on to assess your value?
 - vii. **How were your metrics calculated?** Is there anything we should know about how you calculate your metrics in order to interpret them properly? Do you depart from a 'standard' in any way?

²² The term 'mitigation' here is intended to refer to any techniques and signals that you have relied on to support the relevant advertising use cases in the context of the experiment arm, which are not resulting from the Privacy Sandbox technologies.

- viii. **Which testing mode did you use?** Which testing mode was used and how were data compared between control and treatment?
 - ix. **How long was your experiment?** When was your experiment in the field, and what was the total amount of time it covered? Were there any important dates or events during this time?
 - x. **Did you work together with other market participants?** Did the auctions in your experiments have a similar number of participants as you would expect post TPC deprecation? Do you know if other participants were also experimenting?
 - xi. **What samples sizes did you use and how have you measured uncertainty?** How many auctions make up your treatment and control groups? Did you experience attrition in your samples over time? How have you calculated standard errors (ie ‘margins of error’) for your results?²³
 - xii. **Are there any other aspects of your tests that affect their comprehensiveness?** Are there any other considerations that impact the comprehensiveness of the test? For example, test duration, traffic fraction, geographic scope. As regards geographic scope, we encourage all testers globally to submit their results. We are also be interested in understanding results by geography and results that might be particularly relevant for the UK.
 - xiii. **Do you have any other feedback on testing?** For example, did you face any challenges to testing (eg, in terms of the ease of using Google’s testing environments)?
37. The questions we have listed are designed to help us understand individual (or combined) third-party tests rather than suggest that market participants should depart drastically from what is useful to them. We welcome any feedback on these questions, including on whether market participants require further guidance on how to answer them or why they are important to us.

How to submit your results

38. Any market participants that intend to submit test results to the CMA should plan to do so to privacysandbox@cma.gov.uk by the end of Q2 2024.

²³ As described in paragraph 24, there are a number of methodologies available to calculate standard errors and market participants should choose the most appropriate for their experiment and the metrics they test.

39. The accompanying annex to this note provides examples of tables market participants could use to make their submissions:
- i. Table 1 is a template for submitting results. It is designed to allow market participants to easily and clearly submit an estimate of the impact of the Privacy Sandbox on relevant metrics, alongside estimates of a standard error and confidence interval.²⁴
 - ii. Table 2 is a template for submitting simple descriptive statistics for the metrics market participants use in their testing. It is designed to let us understand the data underlying estimates of impacts.
 - iii. Table 3 is a template for submitting information on the characteristics of treatment and control impressions, described in paragraph 26 above. It is designed to allow the us to straightforwardly understand whether randomisation has been effective and assess the comparability of treatment and control groups.
 - iv. Table 4 is a template for providing answers to questions 36.i-xiii to above. It is designed to allow market participants to easily and clearly provide an answer to each question.
40. Having standardised submissions from market participants will allow us to interpret and understand results quickly. That being said, the templates are intended to illustrate a useful format rather than a strict requirement. If they are not entirely suitable, market participants should view them as basis for their submissions to the CMA.
41. We are under certain statutory obligations to protect confidential information relating to individuals and businesses that comes to it in connection with the exercise of its statutory functions. These statutory obligations include restrictions on the further disclosure of information.
42. Although it may be appropriate that we share the results of testing with Google to facilitate discussions between the CMA and Google, we do not envisage that it would be necessary to identify the source of any such information. Further information on the CMA's approach to disclosure can be found in [CMA6 Transparency and disclosure: Statement of the CMA's policy and approach \(publishing.service.gov.uk\)](#) .

²⁴ As described in paragraph 24, there are a number of methodologies available to calculate standard errors and market participants should choose the most appropriate for their experiment and the metrics they test.

Next steps

43. We are seeking to encourage market participants to engage in testing Google's Privacy Sandbox technologies and submit their results as evidence to inform our assessment at the Standstill Period of the proposals' impact on competition. To facilitate third-party testing, this note has reiterated how market participants can use two experimental designs (as originally proposed in our experiments note of November 2022) to test how the Privacy Sandbox is impacting them, either independently or together with other market participants, as appropriate.
44. In addition, this note has described Google's proposal to enable two testing modes in the Chrome browser starting in Q4 2023 that align with our two proposed experimental designs and will enable market participants to experiment using the same treatment and control groups:
 - i. Mode A will be active from Q4 2023 and aligns with our Design 1 – it will involve the Chrome browser creating treatment and control groups then labelling traffic to let ad techs know to which groups impressions belong.
 - ii. Mode B will be active from Q1 2024 and aligns with our Design 2 – it will involve Chrome deprecating TPC entirely for a slice of traffic.
45. We encourage those market participants that plan on testing to inform the CMA of their plans as soon as they are able to. We welcome any feedback on the experimental designs or Google's testing modes, and clarificatory questions on the questions we have posed about the practical features of tests market participants should answer when submitting results to the CMA (paragraph 36.i-xiii to above).
46. We will continue to engage with market participants in advance of the Privacy Sandbox technologies moving into General Availability in Q3 2023. During this period, our engagement will focus on understanding readiness for testing in the market.
47. Once the technologies move to General Availability, and in particular when Chrome's testing modes become active from Q4 2023, we would like to hear from market participants about their experiences and results from testing. We understand that for many market participants, testing might not be feasible when these testing modes initially become available, but we encourage testing to take place as early as possible.

48. We request that market participants who are able to quantitatively test the Privacy Sandbox (or alternatives as discussed in paragraph 9), do so **before the end of Q2 2024** so we can use them in our assessment.

Annex: Templates for submitting results

1. We suggest that market participants submit their results in four tables. We provide templates for these four tables below:
 - i. Table 1: A table of results. We have included all metrics listed in paragraph 19i of this note. Market participants can exclude any of these metrics from their submission if they do not record them and add other metrics that they deem relevant.
 - ii. Table 2: table of descriptive statistics for the metrics tested and shown in . These statistics will allow us to understand the data underlying estimates. Again, we have included all metrics listed in paragraph 19 of this note, but market participants should exclude those they do not measure, and include others that they do.
 - iii. Table 3 : a table for providing information on the characteristics of treatment and control impressions. We have provided an example of the characteristics we would like market participants to report. However, we understand what can be reported and in what format might vary across participants.
 - iv. Table 4 : A table of answers to our questions in paragraph 36i of this note.
2. As we discussed in paragraph 40, market participants should view these as suggested templates on which to base their submissions rather than a strict requirement. However, if it is suitable to depart from the templates, we request that market participants maintain clarity and ease of interoperability in their formatting.
3. We also encourage market participants to submit a written response with their results providing any necessary description, interpretation, or discussion of their results.

Table 1: template table of experimental results that market participants should submit to the CMA

Metric	Treatment vs control 1	Treatment vs control 2	Control 2 vs control 1
Revenue per impression	<i>effect</i>	<i>effect</i>	<i>effect</i>
	<i>Standard error</i>	<i>Standard error</i>	<i>Standard error</i>
	<i>95% confidence interval</i>	<i>95% confidence interval</i>	<i>95% confidence interval</i>
Clicks per dollar			
Conversions per dollar			
Clicks per impression			
Latency			
Unique viewers reached relative to goal			
% of planned campaign budget spent			

Total unique bid requests served			
Average time spent or completion rates			
Brand lift			

Table 2: template table for reporting descriptive statistics for metrics in the treatment and control groups

Metric	Treatment	Control 1	Control 2
Revenue per impression	<i>Mean</i>	<i>Mean</i>	<i>Mean</i>
	Standard deviation	Standard deviation	Standard deviation
	25 th and 75 th percentile	25 th and 75 th percentile	25 th and 75 th percentile
Clicks per dollar			
Conversions per dollar			
Clicks per impression			
Latency			
% budget spent			
Unique viewers reached relative to goal			
Total unique bid requests served			
Average			

Table 3: template table for reporting characteristics of treatment and control impressions

Characteristic	Treatment	Control 1	Control 2
Country			
%UK			
% EU			
etc....			
<i>Add characteristics as appropriate</i>			

Table 4: template table for answers to the CMA's questions regarding market participants' testing results

Question	Answer. If not applicable, please explain why
Was your testing visible or invisible?	
Which inventory, formats and campaign types did your testing cover?	
How comprehensive were the changes you made for testing to your infrastructure?	
How comprehensive were any mitigations you used?	
How much did you rely on the mitigations you used?	
How business-critical were the metrics you collected?	
How were your metrics calculated?	
Which testing mode did you use?	
How long was your experiment?	
Did you work together with other market participants?	
What samples sizes did you use and how have you measured uncertainty?	
Are there any other aspects of your tests that affect their comprehensiveness?	
Do you have any other feedback on testing?	