Guidehouse
Outwit Complexity

Meta

VRS Compliance Metrics Verification

October 30, 2024

Guidehouse

Table of Contents

Guidehouse

I. Executive Summary

Guidehouse Inc. (Guidehouse or Reviewer) was proposed by Meta Platforms, Inc. (Meta) and has the consent of the United States Department of Justice (DOJ) to serve as the independent third-party Reviewer pursuant to ¶18 of the Settlement Agreement and Final Judgement entered in *United States v. Meta Platforms, Inc*., No. 22-Civ-5187 (S.D.N.Y.) on June 27, 2022, Dkt. No. 7 (Settlement Agreement).[1]

The Reviewer is an independent third-party and, pursuant to Settlement Agreement ¶17, will "review each Compliance Report and verify compliance with the VRS Compliance Metrics."[2]

Pursuant to Settlement Agreement ¶17 and the VRS Compliance Metrics Agreement dated January 6, 2023, Guidehouse reviewed the Meta Compliance Report dated September 30, 2024 for the reporting period from May 1, 2024 to August 31, 2024 (Reporting Period) and verified that Meta complied with the relevant VRS Compliance Metrics for both sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Ad Impressions as well as Housing Advertisements with greater than 1,000 Ad Impressions.[3]

In establishing the VRS Compliance Metrics, Meta's sampling of users to measure the Eligible Audience for the purposes of the VRS Compliance Metrics calculation and use of Differential Privacy (DP) in its implementation of Bayesian Improved Surname Geocoding (BISG) as part of the VRS Compliance Metrics calculation process are included in the VRS Compliance Metrics Agreement dated January 6, 2023.[4] Meta's use of the 50% BISG threshold is discussed in its November 2021 white paper "How Meta is working to assess fairness in relation to race / ethnicity in the U.S. across its products and systems".[5]

For the Reporting Period, Guidehouse verified compliance with the VRS Compliance Metrics by assessing Meta's sampling of Eligible Audience members, implementation of BISG, aggregation of Potential Impressions and Actual Impressions, and computation of Variance and Coverage

---

[1] Capitalized terms are defined in Appendix A – Definitions.

[2] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement ¶17. The Settlement Agreement is available at https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

[3] Meta Platforms, Inc. "VRS Compliance Metrics Agreement." 6 Jan. 2023.

[4] Ibid.

[5] Meta's November 2021 white paper "How Meta is working to assess fairness in relation to race in the U.S. across its products and systems" is available at: https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems.

for accuracy and robustness using synthetic data created by Guidehouse.[6] [7] [8] While certain parameters existed when establishing the VRS Compliance Metrics, Guidehouse reviewed the impact of sampling of Eligible Audience members, DP, and BISG probability thresholds in its analysis of the synthetic data to understand the potential sensitivity of Variance and Coverage to such parameters.

Guidehouse also independently computed Variance, separately for sex and estimated race / ethnicity, for each Housing Advertisement in the Reporting Period using aggregated data provided by Meta. Guidehouse used these Variances to calculate Coverage and compared these calculations to the VRS Compliance Metrics established in the VRS Compliance Metrics Agreement dated January 6, 2023 and to Meta's reported Coverage for the Reporting Period. Guidehouse calculated a difference of zero percent between Meta's Coverage reported in its Compliance Report compared to Guidehouse's independently calculated Coverage across all VRS Compliance Metrics, as shown in Table 1 and Table 2 below. As these values are higher than the required VRS Compliance Metrics, Guidehouse verified Meta's compliance with the VRS Compliance Metrics.

[6] Potential Impressions and Actual Impressions are the field names in the Reporting Period dataset provided by Meta that contain Ad Impressions associated with the Eligible Audience and Actual Audience, respectively.

[7] As disaggregated data from the Reporting Period is not available, Guidehouse created a synthetic dataset to supplement analysis of the Reporting Period data. The synthetic data represented delivery of Housing Advertisements across 10 days.

[8] Guidehouse's implementation of Earth Mover's Distance to calculate Variance is consistent with Meta's implementation, pursuant to the Settlement Agreement and the VRS Compliance Metrics Agreement dated January 6, 2023.

Table 1: Meta's Reported Coverage and Guidehouse's Calculated Coverage for Housing Advertisements with ≥ 300 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage[9] (B)	Guidehouse – Calculated Coverage[10] (C)	Difference in Coverage (C – B)
Sex	≤10%	90.2%	93.7%	93.7%	0.0%
	≤5%	78.3%	86.3%	86.3%	0.0%
Estimated Race / Ethnicity	≤10%	80.1%	83.8%	83.8%	0.0%
	≤5%	56.8%	64.0%	64.0%	0.0%

Table 2: Meta's Reported Coverage and Guidehouse's Calculated Coverage for Housing Advertisements with >1,000 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage[11] (B)	Guidehouse – Calculated Coverage[12] (C)	Difference in Coverage (C – B)
Sex	≤10%	91.7%	95.1%	95.1%	0.0%
	≤5%	84.5%	89.0%	89.0%	0.0%
Estimated Race / Ethnicity	≤10%	81.0%	85.5%	85.5%	0.0%
	≤5%	61.0%	68.8%	68.8%	0.0%

Notwithstanding the verification of Meta's compliance with the VRS Compliance Metrics, Guidehouse had six observations from its analysis of synthetic data and Reporting Period data.

Four observations were based on Guidehouse's analysis of synthetic data and pertained to Meta's sampling of Eligible Audience members to compute the VRS Compliance Metrics, Meta's implementation of DP within BISG, and Meta's selection of the BISG probability threshold.

---

[9] Meta Coverage as reported in Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 – August 31, 2024.

[10] Guidehouse calculations use data aggregated at the Housing Advertisement level provided by Meta for the Reporting Period.

[11] Meta Coverage as reported in Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 – August 31, 2024.

[12] Guidehouse calculations use data aggregated at the Housing Advertisement level provided by Meta for the Reporting Period.

1. **Meta's sampling of users from the Eligible Audience produces a distribution of users that is consistent with random sampling**

   In the synthetic data, Guidehouse found that Meta's sampling of users from the Eligible Audience yields a distribution of sex and estimated race / ethnicity across sampled users that is consistent with random sampling, confirming that Meta's sampling process does not introduce any bias associated with the selection of users into samples.

2. **Variance and Coverage measured for a sample of Eligible Audience members may differ from Variance and Coverage measured for the Eligible Audience**

   Guidehouse found Variance and Coverage calculated for a sample of Eligible Audience members in the synthetic data may differ from the Variance and Coverage calculated for the synthetic Eligible Audience. The VRS Compliance Metrics Agreement dated January 6, 2023 specifies that the Eligible Audience will be measured based on a sample of users. Meta's expected minimum sample size threshold of 4,500 users is sufficient for populations as large as 500 million users, which exceeds the average number of daily Meta platform users and thus is a reasonably sized sample of users.[13] [14] [15] As such, Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics in the Reporting Period is not impacted by this observation.

---

[13] This conclusion is based on a 95% confidence level with a 5% margin of error, which are industry-standard sampling parameters. The confidence level is the probability that the true value being studied falls within a specified range of values. The margin of error denotes the sampling error due to measurement in a sample. https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22.

[14] Meta's sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. The sampling module is designed to raise an alert if more than 1% of the requests to the sampling module within the prior 24 hours delivered a sampled audience of less than 75% of the requested sample size, i.e., an audience of less than 4,500 where the requested size is 6,000 or of less than 75% of the estimated Eligible Audience if it contains fewer than 6,000 total users.

[15] Meta disclosed in its 2023 Annual Report that the average daily active users in the U.S. and Canada ranged from approximately 195 million to 205 million between December 2021 and December 2023. https://www.sec.gov/ixviewer/ix.html?doc=/Archives/edgar/data/0001326801/000132680124000012/meta-20231231.htm.

3. **DP adds noise that may impact Variance and Coverage**

   DP noise added by Meta in its implementation of BISG may impact Variance and Coverage,[16] but the behavior of that noise is not expected to increase Coverage in the Reporting Period data, outside of the following scenarios:

   1. When the concentration of Housing Advertisements immediately above the 5% and 10% Variance thresholds is higher than the concentration immediately below the thresholds; and,
   2. In data sets where the proportion of Housing Advertisements with relatively low counts of Actual Impressions is high.[17]

   In the Reporting Period, the number of Housing Advertisements with Variance immediately below the 5% and 10% Variance thresholds outweighed those with Variance immediately above the thresholds, and, therefore, the first scenario does not impact Guidehouse's verification of compliance with the VRS Compliance Metrics.[18]

   Meta imposes a lower limit of zero to the count of Actual Impressions for a given race / ethnicity bucket for a given Housing Advertisement when DP noise is added, which may result in a decrease in Variance for Housing Advertisements with relatively low counts of Actual Impressions before DP, and thus may increase calculated Coverage. This is consistent with both the definition of Ad Impressions in the Settlement Agreement and with academic and industry literature on the application of DP, and thus is reasonable. As such, this does not impact Guidehouse's verification of compliance with the VRS Compliance Metrics.

4. **Variance and Coverage are sensitive to the BISG probability threshold**

   In the synthetic data, Guidehouse found Variance and Coverage to be sensitive to the probability threshold used in the implementation of BISG. As use of a 50% BISG probability threshold is consistent with academic, industry, and regulatory literature, and thus is reasonable, Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics in the Reporting Period is not impacted by this observation.

---

[16] Neither self-reported nor inferred user race / ethnicity information is maintained in the Meta user database. As such, for the purposes of operating the VRS and calculation of the VRS Compliance Metrics, Meta uses BISG to estimate user race / ethnicity. In its implementation of BISG, Meta applies DP "to prevent adversarial disclosure or re-identification by any party while still enabling aggregate analyses" by adding noise to the aggregated estimated race / ethnicity distributions produced by BISG. Meta's application of privacy enhancement is discussed further in its white papers available at https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems and https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf.

[17] The parameters of DP are defined in the VRS Compliance Metrics Agreement dated January 6, 2023.

[18] See Figure 4 for distribution of Variance observed in the Reporting Period.

Two additional observations were based on Guidehouse's analysis of Reporting Period data and pertained to performance of the Application Programming Interface (API) used in Meta's Eligible Audience sampling process and Ad Impression counts observed in the Reporting Period data.

5. **Meta's Eligible Audience sampling process during the Reporting Period resulted in some Eligible Audience samples that were smaller than the Meta-requested sample size**

   Meta leverages an API to sample users into the Eligible Audience for the purposes of calculating the VRS Compliance Metrics. The sampling API is expected to return at least 75% of the requested number of Eligible Users for at least 99.9% of samples selected, which was exceeded in the Reporting Period.

   In instances where under sampling did occur, Guidehouse assessed the impact of the under sampling on Coverage for cases that resulted in samples containing 385 users or less for at least one of the daily samples selected for a given Housing Advertisement during the Reporting Period.[19] The analysis showed that Meta exceeded the VRS Compliance Metrics in the Reporting Period by more than the computed potential overestimation of Coverage for both the 5% and 10% Variance thresholds for sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Impressions and Housing Advertisements with more than 1,000 Impressions. As the margin by which Meta exceeded the VRS Compliance Metrics in the Reporting Period is larger than the calculated potential impact of the under sampling, the Eligible Audience sampling observation does not impact Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics for the Reporting Period.

[19] 385 users is the minimum sample size required to ensure a 95% confidence level with a 5% margin of error for populations as large as 500 million users, which exceeds Meta's daily average usage in the United States. 95% confidence and 5% margin of error are industry-standard sampling parameters. The confidence level is the probability that the true value being studied falls within a specified range of values. The margin of error denotes the sampling error due to measurement in a sample. See https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22 for sample size calculations and https://www.sec.gov/ixviewer/ix.html?doc=/Archives/edgar/data/0001326801/000132680124000012/meta-20231231.htm for Meta Annual Report's disclosure of average daily platform usage.

6. **Meta's decisions related to the treatment of unknown ZIP Code Tabulation Areas (ZCTA),[20] ZCTAs with low populations, Housing Advertisements with small daily Audiences, and unknown sex may result in a subset of Ad Impressions not being captured in VRS Compliance Metrics calculations**

   Guidehouse noted differences in Ad Impression counts for a given Housing Advertisement when Ad Impressions are counted across sex versus across estimated race / ethnicity. The discrepancies noted are due to Meta's treatment of unknown ZCTAs or sex, ZCTAs with populations too small for BISG to accurately estimate race / ethnicity, and Housing Advertisements with Eligible Audiences or Actual Audiences that are not large enough to implement DP, which may result in some Ad Impressions being omitted from the calculation of Variance and Coverage. The collective impact of these omissions was not large enough to affect Coverage in the Reporting Period and, therefore, Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics in the Reporting Period is not impacted by this observation.

When considering the maximum potential impact of these observations from the analysis of Reporting Period data in aggregate, the maximum potential impact on Coverage computed was smaller than the margin by which Meta exceeded the VRS Compliance Metrics in the Reporting Period.[21] Therefore, these observations, in aggregate, do not impact Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics for the Reporting Period.

[20] Terminology has been updated to replace references to ZIP Codes with references to ZIP Code Tabulation Areas to ensure consistency with the VRS Compliance Metrics Agreement dated January 6, 2023 and with Meta's and Guidehouse's processing of the synthetic and Reporting Period data. There is no change to the underlying data or analyses performed for either the synthetic data analysis or the actual Reporting Period data, which have always leveraged ZCTAs.

[21] Observations include both impacts of system functionality in the VRS Compliance Metrics calculation process and impacts of certain users not providing sex data or Meta not having sufficient information to estimate a user's ZCTA.

Guidehouse

II. Verification of VRS Compliance Metrics

For the Reporting Period, Guidehouse verified that Meta complied with the VRS Compliance Metrics for both sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Ad Impressions as well as Housing Advertisements with greater than 1,000 Ad Impressions, in accordance with the Settlement Agreement and the VRS Compliance Metrics Agreement dated January 6, 2023.

Guidehouse independently computed Variance, separately for sex and estimated race / ethnicity, for each Housing Advertisement in the Reporting Period using aggregated data provided by Meta. Guidehouse used these Variances to calculate Coverage and compared such calculations to the VRS Compliance Metrics established in the VRS Compliance Metrics Agreement dated January 6, 2023 and to Meta's reported Coverage for the Reporting Period.

In Table 3 and Table 4 below, Guidehouse summarized the target Coverage at the agreed upon Variance thresholds for sex and estimated race / ethnicity for the Reporting Period, along with Meta's Coverage reported in its Compliance Report compared to Guidehouse's independently calculated Coverage.[22] The difference in Coverage computed by Meta and computed by Guidehouse across all VRS Compliance Metrics was zero percent, and these figures were higher than the required VRS Compliance Metrics.

Table 3: Meta's Reported Coverage and Guidehouse's Calculated Coverage for Housing Advertisements with ≥ 300 Ad Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage[23] (B)	Guidehouse – Calculated Coverage[24] (C)	Difference in Coverage (C – B)
Sex	≤10%	90.2%	93.7%	93.7%	0.0%
	≤5%	78.3%	86.3%	86.3%	0.0%
Estimated Race / Ethnicity	≤10%	80.1%	83.8%	83.8%	0.0%
	≤5%	56.8%	64.0%	64.0%	0.0%

[22] Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 - August 31, 2024.

[23] Meta Coverage as reported in Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 - August 31, 2024.

[24] Guidehouse calculations use data aggregated at the Housing Advertisement level provided by Meta for the Reporting Period.

Table 4: Meta's Reported Coverage and Guidehouse's Calculated Coverage for Housing Advertisements with >1,000 Ad Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage[25] (B)	Guidehouse – Calculated Coverage[26] (C)	Difference in Coverage (C – B)
Sex	≤10%	91.7%	95.1%	95.1%	0.0%
	≤5%	84.5%	89.0%	89.0%	0.0%
Estimated Race / Ethnicity	≤10%	81.0%	85.5%	85.5%	0.0%
	≤5%	61.0%	68.8%	68.8%	0.0%

[25] Meta Coverage as reported in Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 - August 31, 2024.

[26] Guidehouse calculations use data aggregated at the Housing Advertisement level provided by Meta for the Reporting Period.

III. Observations

While verifying the VRS Compliance Metrics for the Reporting Period, Guidehouse made six observations, four based on its analysis of the synthetic data and two based on its analysis of the Reporting Period data.

1. Observations from review of synthetic data

a. Meta's sampling of users from the Eligible Audience produces a distribution of users that is consistent with random sampling

Conclusion:

Guidehouse observed that Meta's sampling of users from the Eligible Audience yields a distribution of sex and estimated race / ethnicity across sampled users that is consistent with random sampling, confirming that Meta's sampling process does not introduce any bias associated with the selection of users into samples.

Supporting Analysis:

To assess whether Meta's sampling process yields a similar user distribution across sex and estimated race / ethnicity as a randomly selected sample, Guidehouse created a synthetic dataset containing Ad IDs and User IDs and compared the distribution of users across sex and estimated race / ethnicity in Meta-selected samples and Guidehouse-selected samples of synthetic users.[27] [28]

The synthetic data had, on average, between 11,968 and 89,966 daily users per synthetic Housing Advertisement. Meta selected 30 samples of synthetic users for each synthetic Housing Advertisement across each of the 10 days in the synthetic data. Meta deployed the same sampling process used in the VRS Compliance Metrics calculation process when sampling from the synthetic data.[29] Meta's sample sizes ranged between 5,680 and 6,000

---

[27] The VRS Compliance Metrics Agreement dated January 6, 2023 establishes that, for the purposes of measuring the Impression distribution across sex and estimated race / ethnicity, Meta selects a sample of users from the Eligible Audience for each Housing Advertisement that fits the targeting options selected by the advertiser and that the sampling process approximates a random sample. For a given Housing Advertisement, the Eligible Audience exists ephemerally before being sampled due to data storage limitations.

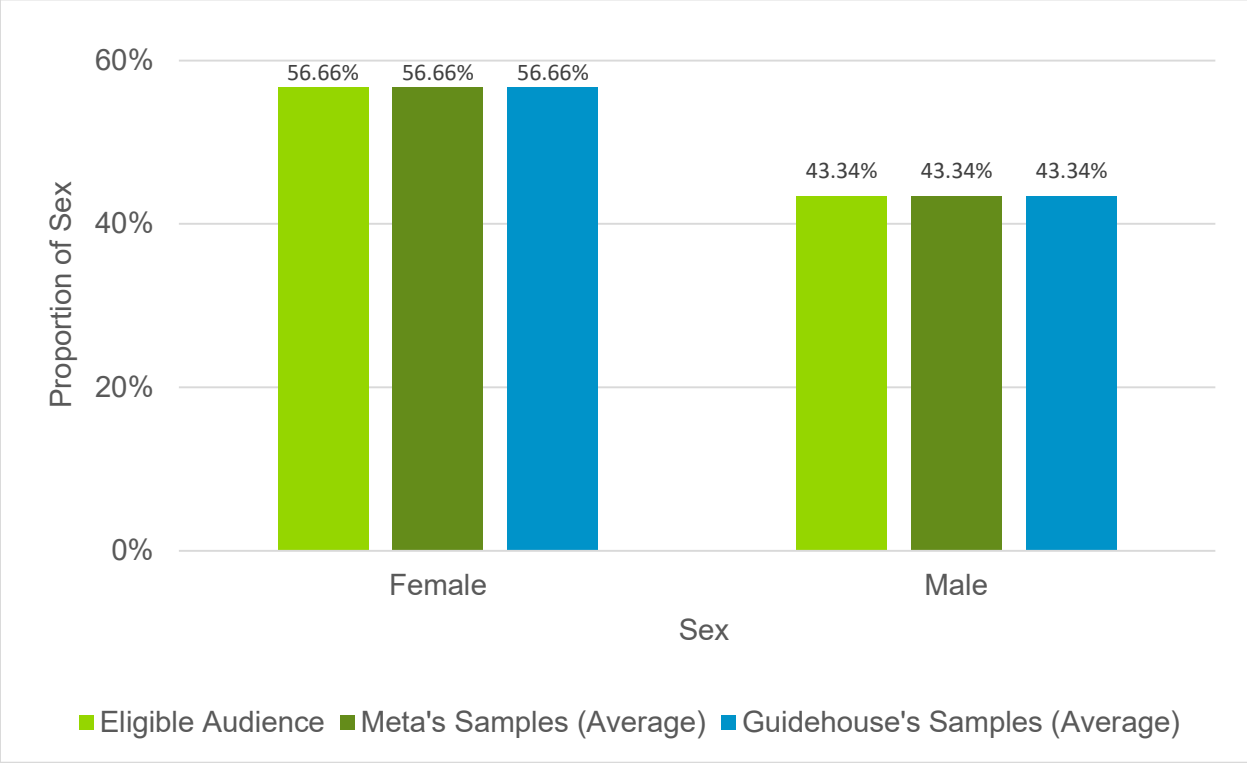
[28] Meta represented that its sampling process in production relies only on Ad IDs and hashed Meta User IDs and, therefore, does not consider demographic characteristics or Impression data that are used for the computation of Variance and Coverage.

[29] Meta's sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. The sampling module is designed to raise an alert if more than 1% of the requests to the sampling module within the prior 24 hours delivered a sampled audience of less than 75% of the requested sample size, i.e., an audience of less than 4,500 where the requested size is 6,000 or of less than 75% of the estimated Eligible Audience if it contains fewer than 6,000 total users.

users, with an average of 5,966 daily users. Guidehouse's sample size was consistently 6,000 users for all Housing Advertisements.

Figure 1 below demonstrates the breakdown of synthetic users by sex in the synthetic Eligible Audience data, Meta's samples, and Guidehouse's samples. The percentage associated with each sex is consistent in the synthetic Eligible Audience dataset and in both samples. For example, the proportions of synthetic users that are female are approximately 56.66% of each of the Eligible Audience population, Meta's samples, and Guidehouse's samples. The differences between the breakdown of synthetic users by sex in the synthetic Eligible Audience data, Meta's samples, and Guidehouse's samples are not statistically significant. [30] [31]

Figure 1. Breakdown of Synthetic Users by Sex



[30] The differences in the proportions of female users between the Eligible Audience data and the samples generated by Meta and Guidehouse are approximately 0.002% in absolute value, and, therefore, are not easily identifiable in Figure 1.

[31] The differences are not statistically significant at the 5% level. 5% statistical significance level is the most commonly used level in hypothesis testing. See https://www.sciencedirect.com/topics/mathematics/significance-level-alpha#:~:text=The%205%20percent%20level%20of,0%20when%20it%20is%20true.

Similarly, Figure 2 below demonstrates the breakdown of synthetic users by estimated race / ethnicity in the synthetic Eligible Audience data, Meta's samples, and Guidehouse's samples. The percentage associated with each estimated race / ethnicity is consistent in the synthetic Eligible Audience dataset and in both samples. For example, the proportions of synthetic users that are categorized as "White" are approximately 56.29% of both the Eligible Audience population and Guidehouse's samples and 56.31% in Meta's samples. The differences between the breakdown of synthetic users by estimated race / ethnicity in the synthetic Eligible Audience data, Meta's samples, and Guidehouse's samples are not statistically significant. [32] [33]

Figure 2. Breakdown of Synthetic Users by Estimated Race / Ethnicity

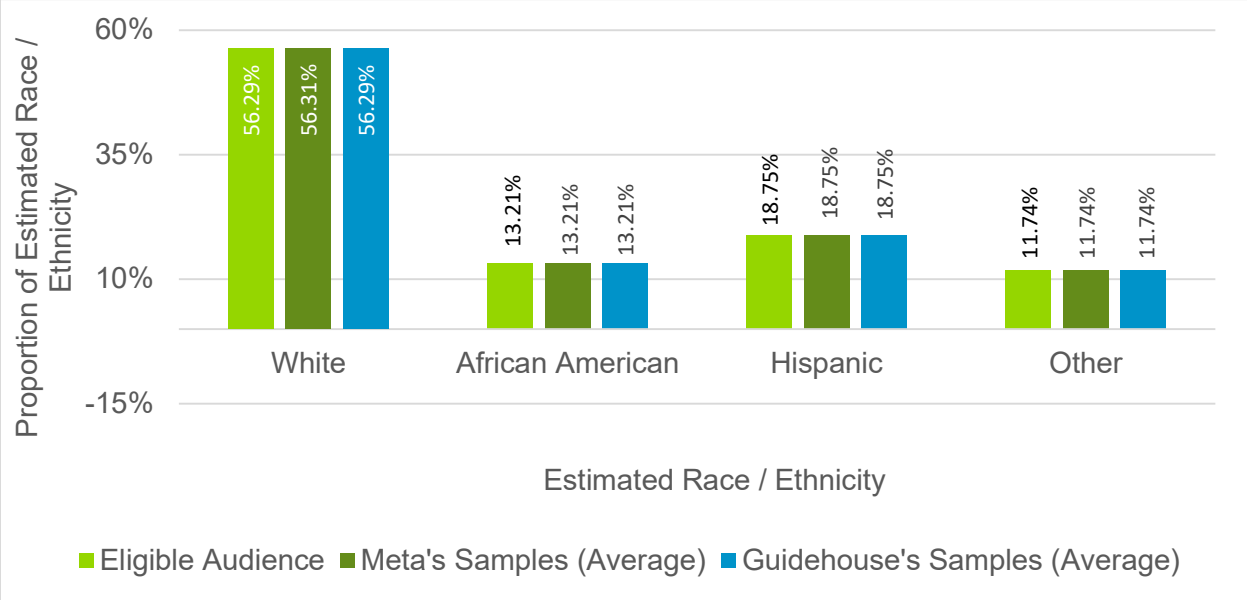


Figure 1 and Figure 2 show that Meta's sampling of users from the Eligible Audience yields a distribution of sex and estimated race / ethnicity across sampled users that is consistent with random sampling, confirming that Meta's sampling process does not introduce any bias associated with the selection of users into samples.

---

[32] The differences in the proportions of "White" users between the Eligible Audience data and the samples generated by Meta and Guidehouse are approximately 0.02% in absolute value, and, therefore, are not easily identifiable in Figure 2.

[33] The differences are not statistically significant at the 5% level. 5% statistical significance level is the most commonly used level in hypothesis testing. See https://www.sciencedirect.com/topics/mathematics/significance-level-alpha#:~:text=The%205%20percent%20level%20of,0%20when%20it%20is%20true.

b. Variance and Coverage measured for a sample of Eligible Audience members may differ from Variance and Coverage measured for the Eligible Audience

Conclusion:

To the extent that the distribution of users across sex and / or estimated race / ethnicity vary from the distribution of Eligible Ad Impressions across the demographic characteristics, sampling of Eligible Audience members may impact the measurement of Variance and Coverage. While this analysis shows that the Coverage evaluated at both the 5% and 10% Variance thresholds in samples of the Eligible Audience may differ from the Coverage calculated for the Eligible Audience, the VRS Compliance Metrics Agreement dated January 6, 2023 establishes that a sample of users will be used for the purposes of measuring the Eligible Audience and calculating Variance and Coverage.

Meta's production system has a target sample size of 6,000 users with an expected minimum sample size threshold of 4,500 users for Housing Advertisements with at least 6,000 users in the Eligible Audience. A sample size of this magnitude is sufficient for populations as large as 500 million users, which exceeds the average number of daily Meta platform users and thus is a reasonably sized sample of users.[34] [35]

Supporting Analysis:

Guidehouse performed an analysis of the Variance and Coverage metrics for sex and estimated race / ethnicity in both Meta's and Guidehouse's samples and compared the metrics against those calculated in the Eligible Audience. We summarize the results of the analysis for Coverage below.

---

[34] This conclusion is based on a minimum sample size of 385 users required to ensure a 95% confidence level with a 5% margin of error for populations as large as 500 million users, which exceeds Meta's daily average usage in the United States. 95% confidence and 5% margin of error are industry-standard sampling parameters. The confidence level is the probability that the true value being studied falls within a specified range of values. The margin of error denotes the sampling error due to measurement in a sample. See https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22 for sample size calculations and https://www.sec.gov/ixviewer/ix.html?doc=/Archives/edgar/data/0001326801/000132680124000012/meta-20231231.htm for Meta Annual Report's disclosure of average daily platform usage. For samples that contained fewer than 385 users in the Reporting Period, Guidehouse assessed the impact on Coverage in Section III.2.b – Observations from review of Reporting Period data.

[35] Meta's sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. The sampling module is designed to raise an alert if more than 1% of the requests to the sampling module within the prior 24 hours delivered a sampled audience of less than 75% of the requested sample size, i.e., an audience of less than 4,500 where the requested size is 6,000 or of less than 75% of the estimated Eligible Audience if it contains fewer than 6,000 total users. In the Reporting Period, the API delivered a sample exceeding the expected minimum sample size more than 99.9% of the time.

For both sex and estimated race / ethnicity, Guidehouse tested separately whether Meta's average Coverage and Guidehouse's average Coverage across the 30 samples were statistically different than the Eligible Audience Coverage and found that average Coverage across both Meta's and Guidehouse's 30 Eligible Audience samples may be statistically different from the Coverage observed in the full synthetic Eligible Audience.[36]

Table 5A below demonstrates that, at the 5% Variance threshold, the Coverage for sex in the Eligible Audience, Meta's samples, and Guidehouse's samples was 79.43%, 78.85%, and 79.32%, respectively, for synthetic Housing Advertisements with at least 300 Impressions in the synthetic dataset. Coverage for estimated race / ethnicity in the Eligible Audience, Meta's samples, and Guidehouse's samples at the 5% Variance threshold for synthetic Housing Advertisements with at least 300 Impressions was 53.48%, 52.87%, and 53.46%, respectively.

Table 5A: Comparison of Coverage for Sex and Estimated Race / Ethnicity Across the Eligible Audience and Meta's and Guidehouse's Samples at the 5% Variance Threshold (Housing Advertisements with ≥ 300 Impressions)

	Sex	Estimated Race / Ethnicity
Eligible Audience	79.43%	53.48%
Meta's Samples*	78.85%	52.87%
Guidehouse's Sample*	79.32%	53.46%

**Average across 30 samples generated by Meta and Guidehouse separately*

For sex, Meta's samples underestimated the Eligible Audience Coverage by, on average, 0.58% (79.43% - 78.85%) and Guidehouse's samples underestimated the Eligible Audience Coverage by, on average, 0.11% (79.43% - 79.32%), where both differences were statistically significant at the 5% level. For estimated race / ethnicity, Meta's samples underestimated Eligible Audience Coverage by, on average, 0.61% (53.48% - 52.87%) and Guidehouse's samples underestimated Eligible Audience Coverage by, on average, 0.02% (53.48% - 53.46%), where both differences were statistically significant at the 5% level.[37]

Table 5B below demonstrates that, at the 5% Variance threshold, the Coverage for sex in the Eligible Audience, Meta's samples, and Guidehouse's samples were 79.48%, 78.96%, and 79.37%, respectively, for synthetic Housing Advertisements with greater than 1,000 Impressions in the synthetic dataset. Coverage for estimated race / ethnicity in the Eligible Audience, Meta's samples, and Guidehouse's samples at the 5% Variance threshold for synthetic Housing

[36] Results were evaluated at the 5% statistical significance level. 5% statistical significance level is the most commonly used level in hypothesis testing. See https://www.sciencedirect.com/topics/mathematics/significance-level-alpha#:~:text=The%205%20percent%20level%20of,0%20when%20it%20is%20true.

[37] Ibid.

Advertisements with greater than 1,000 Impressions were 56.62%, 56.02%, and 56.62%, respectively.

Table 5B: Comparison of Coverage for Sex and Estimated Race / Ethnicity Across the Eligible Audience and Meta's and Guidehouse's Samples at the 5% Variance Threshold (Housing Advertisements with >1,000 Impressions)

	Sex	Estimated Race / Ethnicity
Eligible Audience	79.48%	56.62%
Meta's Samples*	78.96%	56.02%
Guidehouse's Sample*	79.37%	56.62%

**Average across 30 samples generated by Meta and Guidehouse separately*

Meta's average Coverage for sex and estimated race / ethnicity were lower by 0.52% (79.48%-78.96%) and 0.60% (56.62%-56.02%), respectively, than the Eligible Audience Coverage, resulting in statistically significant differences at the 5% level. Similarly, Guidehouse's average Coverage for sex was lower by 0.11% (79.48%-79.37%), where such difference was statistically significant at the 5% level. However, Guidehouse's average Coverage for estimated race / ethnicity resulted in a difference of 0.00% (56.62%-56.62%), compared to the Eligible Audience Coverage. [38]

Guidehouse also evaluated the impact of sampling on Coverage at the 10% Variance threshold. Table 6A below demonstrates that, at the 10% Variance threshold, the Coverage for sex in the Eligible Audience, Meta's samples, and Guidehouse's samples was 97.88%, 97.83%, and 97.87%, respectively, for Housing Advertisements with at least 300 Impressions in the synthetic dataset. Coverage for estimated race / ethnicity in the Eligible Audience, Meta's samples, and Guidehouse's samples at the 10% Variance threshold for synthetic Housing Advertisements with at least 300 Impressions was 78.05%, 78.05%, and 78.07%, respectively.

Table 6A: Comparison of Coverage for Sex and Estimated Race / Ethnicity Across the Eligible Audience and Meta's and Guidehouse's Samples at the 10% Variance Threshold (Ads with ≥ 300 Impressions)

	Sex	Estimated Race / Ethnicity
Eligible Audience	97.88%	78.05%
Meta's Samples*	97.83%	78.05%
Guidehouse's Sample*	97.87%	78.07%

**Average across 30 samples generated by Meta and Guidehouse separately*

[38] Ibid.

For sex, Meta's samples underestimated the Eligible Audience Coverage by, on average, 0.05% (97.88%-97.83%) and Guidehouse's samples underestimated the Eligible Audience Coverage by, on average, 0.01% (98.88%-97.87%), where both differences were statistically significant at the 5% level. For estimated race / ethnicity, Meta's samples produced an average Coverage that is indifferent from the Eligible Audience Coverage, resulting in a difference of 0.00% (78.05%-78.05%), and Guidehouse's samples overestimated Eligible Audience Coverage by, on average, 0.02% (78.05%-78.07%), however, such difference was not statistically significant at the 5% level.[39]

Table 6B below demonstrates that, at the 10% Variance threshold, the Coverage for sex in the Eligible Audience, Meta's samples, and Guidehouse's samples was 97.88%, 97.81%, and 97.87%, respectively, for synthetic Housing Advertisements with greater than 1,000 Impressions in the synthetic dataset. Coverage for estimated race / ethnicity in the Eligible Audience, Meta's samples, and Guidehouse's samples at the 10% Variance threshold for synthetic Housing Advertisements with greater than 1,000 Impressions was 78.27%, 78.29%, and 78.28%, respectively.

Table 6B: Comparison of Coverage for Sex and Estimated Race / Ethnicity Across the Eligible Audience and Meta's and Guidehouse's Samples at the 10% Variance Threshold (Housing Advertisements with >1,000 Impressions)

	Sex	Estimated Race / Ethnicity
Eligible Audience	97.88%	78.27%
Meta's Samples*	97.81%	78.29%
Guidehouse's Sample*	97.87%	78.28%

**Average across 30 samples generated by Meta and Guidehouse separately*

For sex, Meta's samples underestimated the Eligible Audience Coverage by, on average, 0.07% (97.88% - 98.81%), where such difference was statistically significant at the 5% level. Guidehouse's samples underestimated the Eligible Audience Coverage by, on average, 0.01% (97.88% - 97.87%), where such difference was not statistically significant at the 5% level. For estimated race / ethnicity, Meta's samples and Guidehouse's samples overestimated Eligible Audience Coverage by, on average, 0.02% (78.27% - 78.29%) and 0.01% (78.27%-78.28%), respectively, where neither difference was statistically significant at the 5% level.[40]

While this analysis shows that the Coverage evaluated at both the 5% and 10% Variance thresholds in samples of the Eligible Audience may differ from the Coverage calculated for the Eligible Audience, the VRS Compliance Metrics Agreement dated January 6, 2023 establishes

[39] Ibid.

[40] Ibid.

that a sample of users will be used for the purposes of measuring the Eligible Audience and calculating Variance and Coverage. Meta's expected minimum sample size threshold of 4,500 users for Housing Advertisements with at least 6,000 users in the Eligible Audience in production is a sample size sufficient for populations as large as 500 million users, which exceeds the average number of daily Meta platform users and thus is a reasonably sized sample of users. [41] [42]

c. DP adds noise that may impact Variance and Coverage

Conclusion:

DP noise added by Meta in its implementation of BISG may impact Variance and Coverage, [43] but the behavior of that noise is not expected to increase Coverage in the Reporting Period data, outside of the following scenarios:

---

[41] This conclusion is based on a minimum sample size of 385 users required to ensure a 95% confidence level with a 5% margin of error, for populations as large as 500 million users, which exceeds Meta's daily average usage in the United States. 95% confidence and 5% margin of error are industry-standard sampling parameters. The confidence level is the probability that the true value being studied falls within a specified range of values. The margin of error denotes the sampling error due to measurement in a sample. See https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22 for sample size calculations and https://www.sec.gov/ixviewer/ix.html?doc=/Archives/edgar/data/0001326801/000132680124000012/meta-20231231.htm for Meta Annual Report's disclosure of average daily platform usage. For samples that contained fewer than 385 users in the Reporting Period, Guidehouse assessed the impact on Coverage in Section III.2.b – Observations from review of Reporting Period data below.

[42] Meta's sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. The sampling module is designed to raise an alert if more than 1% of the requests to the sampling module within the prior 24 hours delivered a sampled audience of less than 75% of the requested sample size, i.e., an audience of less than 4,500 where the requested size is 6,000 or of less than 75% of the estimated Eligible Audience if it contains fewer than 6,000 total users. In the Reporting Period, the API delivered a sample exceeding the expected minimum sample size more than 99.9% of the time.

[43] Neither self-reported nor inferred user race / ethnicity information is maintained in the Meta user database. As such, for the purposes of operating the VRS and calculation of the VRS Compliance Metrics, Meta uses BISG to estimate user race / ethnicity. In its implementation of BISG, Meta applies DP "to prevent adversarial disclosure or re-identification by any party while still enabling aggregate analyses" by adding noise to the aggregated estimated race / ethnicity distributions produced by BISG. Meta's application of privacy enhancement is discussed further in its white papers available at https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems and https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf.

1. When the concentration of Housing Advertisements immediately above the 5% and 10% Variance thresholds is higher than the concentration immediately below the thresholds; and,
2. In data sets where the proportion of Housing Advertisements with relatively low counts of Actual Impressions is high.[44]

In the Reporting Period, the number of Housing Advertisements with Variance immediately below the 5% and 10% Variance thresholds outweighed those with Variance immediately above the thresholds, thus the first scenario does not impact Guidehouse's verification of compliance with the VRS Compliance Metrics.[45]

Meta imposes a lower limit of zero to the count of Actual Impressions for a given race / ethnicity bucket for a given Housing Advertisement when DP noise is added, which may result in a decrease in Variance for Housing Advertisements with relatively low counts of Actual Impressions before DP, and thus may increase calculated Coverage. This is consistent with both the definition of Ad Impressions in the Settlement Agreement and with academic and industry literature on the application of DP, and thus is reasonable. As such, this does not impact Guidehouse's verification of compliance with the VRS Compliance Metrics.

Supporting Analysis:

To evaluate the impact of DP on Variance and Coverage, Guidehouse generated synthetic user and Housing Advertisement data, representing Housing Advertisement delivery across ten days, and compared the results of Meta's processing of the synthetic data, which included the addition of DP, to the results of Guidehouse's processing of the synthetic data, which did not include DP.[46] [47] Meta processed the synthetic data 30 times, which produced 30 distinct sets of aggregated estimated race / ethnicity, Variance, and Coverage for the synthetic data. For the analysis, Guidehouse calculated the average Variance across Meta's 30 runs for each Housing Advertisement and assigned the average Variance to that Housing Advertisement, to enable comparisons across Meta's and Guidehouse's Variance distributions.

---

[44] The parameters of DP are defined in the VRS Compliance Metrics Agreement dated January 6, 2023.

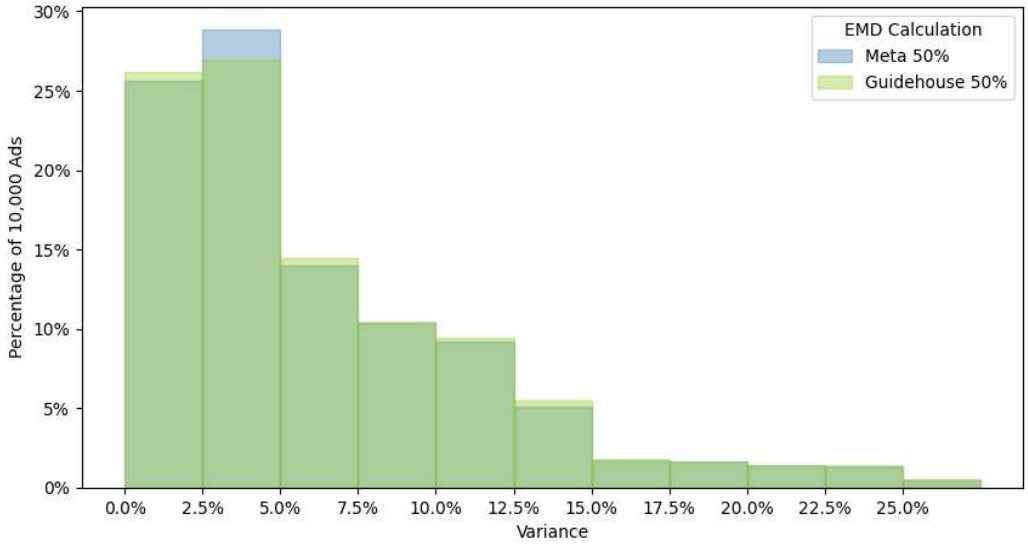
[45] See Figure 4 for distribution of Variance observed in the Reporting Period.

[46] As another privacy measure, Meta does not calculate impressions totals across estimated race / ethnicity via BISG on a given day for a Housing Advertisement if the number of users and Actual Impressions is below a predefined threshold. Instead, Meta "rolls over" those Ad Impressions to the next day for BISG inference with DP. Guidehouse tested this by simulating daily Housing Advertisement delivery of synthetic Housing Advertisements with users and Actual Impressions below Meta's threshold for specific synthetic Housing Advertisements on certain days to confirm Actual Impressions were rolled over to the subsequent day.

[47] Meta's and Guidehouse's implementation of BISG with a 50% probability threshold did not produce statistically significant differences in results. Therefore, race / ethnicity estimated through BISG, aggregation of the data, and computation of Variance and Coverage were sufficiently similar in this analysis to isolate the impact of DP.

Figure 3 below provides a comparison of the distribution of average Variance generated by Meta and the distribution of Variance generated by Guidehouse for all Housing Advertisements in the synthetic data.

Figure 3. Comparison of Meta's (with DP) and Guidehouse's (without DP) Variance Distribution for Estimated Race / Ethnicity



The average Variance computed by Meta across all Housing Advertisements in the synthetic data was 6.37%, versus an average Variance of 6.41% computed by Guidehouse. The minimum and maximum average Variance calculated by Meta was 0.23% and 33.42%, respectively, as compared to 0.09% and 33.45% computed by Guidehouse.

To provide further insight regarding the impact of DP on Variance, Guidehouse analyzed the fluctuation in the Variance computed by Meta for each Housing Advertisement across its 30 runs of BISG. Guidehouse observed the magnitude of the impact of DP on Variance differed across the 30 runs, despite consistent underlying impression data. The Housing Advertisement with the smallest observed fluctuation in Variance across the 30 runs had a minimum computed Variance of 8.25% and a maximum computed Variance of 8.29%, or a spread of 0.04%. The Housing Advertisement with the largest observed fluctuation in Variance had a minimum computed Variance of 3.12% and a maximum computed Variance of 14.07%, or a spread of 10.95%. These results indicate that the magnitude of the potential impact of DP on Variance may fluctuate.

In this analysis using synthetic data, the differences between Meta's and Guidehouse's Variance computations also resulted in discrepancies in the Coverage, as demonstrated in Table 7A for synthetic Housing Advertisements with at least 300 Ad Impressions and in Table 7B for synthetic Housing Advertisements with more than 1,000 Ad Impressions.

Table 7A: Comparison of Meta's (with DP) and Guidehouse's (without DP) Variance and Coverage (Housing Advertisements with ≥ 300 Impressions)

	Meta*	Guidehouse	Difference
Variance	6.37%	6.41%	-0.04%
Coverage at Variance <= 5%	54.06%	53.48%	0.58%
Coverage at Variance <= 10%	78.61%	78.05%	0.56%

**Average across all of Meta's 30 runs*

As Table 7A shows, Guidehouse's computed Coverage for the synthetic data at the 5% Variance threshold was 53.48%, compared to Meta's average Coverage of 54.06% based on the average Variance across 30 runs.[48] Therefore, at the 5% Variance threshold, the difference in Variance caused a difference of 0.58% in Coverage between Guidehouse's calculation and Meta's average calculation. At the 10% Variance threshold, Guidehouse's computed Coverage was 78.05% compared to Meta's average Coverage of 78.61% based on the average Variance across 30 runs, resulting in a difference of 0.56% in Coverage between Guidehouse's calculation and Meta's average calculation.

Table 7B: Comparison of Meta's (with DP) and Guidehouse's (without DP) Variance and Coverage (Housing Advertisements with > 1,000 Impressions)

	Meta*	Guidehouse	Difference
Variance	6.24%	6.24%	0.00%
Coverage at Variance <= 5%	56.80%	56.62%	0.18%
Coverage at Variance <= 10%	78.51%	78.27%	0.24%

**Average across all of Meta's 30 runs*

As Table 7B shows, Guidehouse's computed Coverage for the synthetic data at the 5% Variance threshold was 56.62%, compared to Meta's average Coverage of 56.80% across 30 runs.[49] Therefore, at the 5% Variance threshold, the difference in Variance caused a difference of 0.18% in Coverage between Guidehouse's calculation and Meta's average calculation. At the 10% Variance threshold, Guidehouse's computed Coverage was 78.27% compared to Meta's average Coverage of 78.51% across 30 runs, resulting in a difference of 0.24% in Coverage between Guidehouse's calculation and Meta's average calculation.

[48] Meta's Coverage across 30 runs ranged between 53.87% and 54.30% at the 5% Variance threshold and between 78.35% and 78.84% at the 10% Variance threshold.

[49] Meta's Coverage across 30 runs ranged between 56.44% and 57.16% at the 5% Variance threshold and between 78.33% and 78.76% at the 10% Variance threshold.

Based on these results, Guidehouse observed that DP may have an impact on the computed Variance and Coverage, and that the impact may fluctuate. To the extent that DP creates a bias in the distribution of impressions, the magnitude and direction of this bias may lead to changes in Coverage.

In previous reporting periods, Meta provided a mathematical explanation of the behavior of DP noise, which posits that the effect of the noise on calculated Variance is inversely related to the difference between the Potential Impression distribution and Actual Impression distribution. Therefore, the effect of the DP noise on Variance is expected to be larger for smaller differences in the distributions, and smaller for larger differences. Meta also analyzed the impact of DP across 100 distinct implementations for both synthetic data and Meta Housing Advertisement data, which provided empirical evidence that the average noise resulting from DP increased Variance.[50] Through further interrogation of the application of DP in the synthetic data, Guidehouse identified two scenarios where the expected impact of DP on Variance and Coverage may not hold, which are discussed further below. The first scenario was identified in the analysis performed in conjunction with the VRS Compliance Metrics Verification Report issued on October 30, 2023 and the second was identified in the analysis performed in conjunction with the current VRS Compliance Metrics Verification Report. The scenarios identified are as follows:

1. When the concentration of Housing Advertisements immediately above the 5% and 10% Variance thresholds is higher than the concentration immediately below the thresholds; and,
2. In data sets where the proportion of Housing Advertisements with relatively low counts of Actual Impressions is high.[51]

<u>Scenario 1:</u> There is a higher concentration of Housing Advertisements immediately above the Variance threshold than immediately below:

While the expected value of Variance with DP is higher than Variance without DP, the application of DP may result in lower Variance for a given Housing Advertisement, as the lower bound of the potential distribution of Variance with DP may be less than the Variance without DP. If values of Variance without the addition of DP noise across Housing Advertisements in the population are clustered immediately above the 5% or 10% Variance thresholds, there may be an increase in Coverage when DP is applied. However, if the proportion of Housing

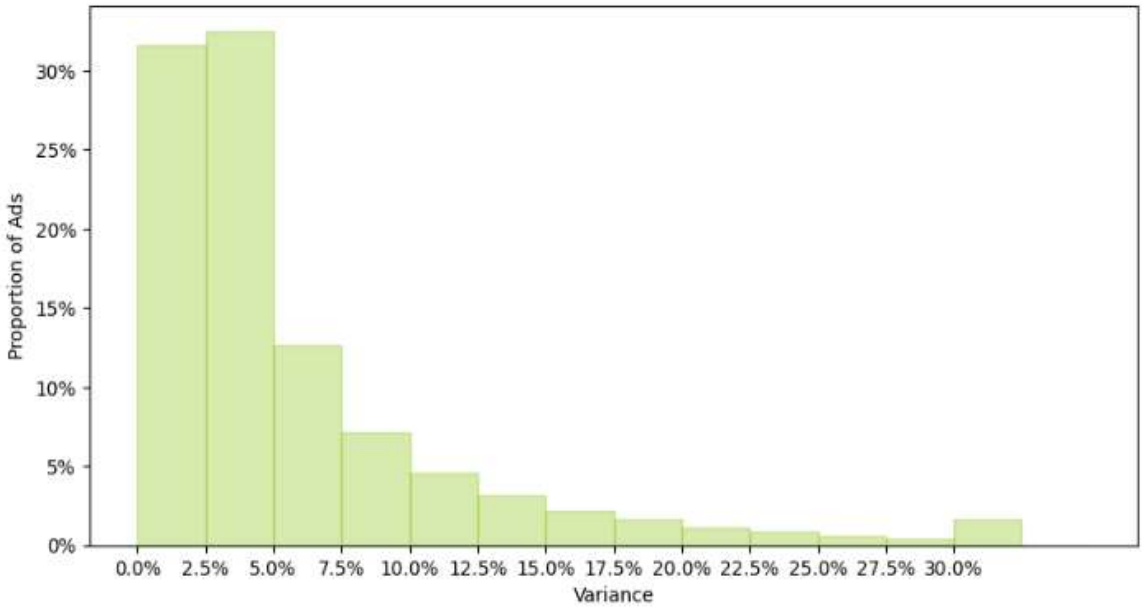
[50] Meta's analysis consisted of first adding DP noise to Potential Impression distributions and Actual Impression distributions for Housing Advertisements in both the synthetic data and a sample of Housing Advertisements from Meta data and computing Variance for each Advertisement. Meta assumed this computed Variance to be the value of Variance without additive noise for each Advertisement. Meta then added DP noise one additional time to the assumed value of Variance without additive noise for each Advertisement and calculated the average difference in Variance between the second application of DP and the assumed value of Variance without additive noise for the 100 runs.

[51] The parameters of DP are defined in the VRS Compliance Metrics Agreement dated January 6, 2023.

Advertisements with Variance immediately below a threshold is higher than those with Variance immediately above the threshold, the impact on DP will result in a decrease in Coverage, on average.

Figure 4 below shows the distribution of Variance for Housing Advertisements in Meta's Reporting Period data.

Figure 4: Distribution of Variance for Estimated Race / Ethnicity for Housing Advertisements in Meta's Fifth Reporting Period Data



As shown in Figure 4, Variance observed in Meta's Reporting Period data is clustered below both the 5% and 10% Variance thresholds. Further, for both the 5% and 10% Variance thresholds, the number of Housing Advertisements with observed Variance immediately below the threshold outweighed the number of Housing Advertisements with observed Variance immediately above the threshold. As such, the impact of DP in aggregate will result in a calculated Coverage at or below the Coverage without DP applied for the Reporting Period data.

<u>Scenario 2:</u> The proportion of Housing Advertisements in the data set with a low count of Actual Impressions is relatively high: [52]

In their implementation of BISG, Meta imposes a system limitation on the noise added to Impression counts by DP to ensure that the addition of DP noise never yields a count of Impressions for given estimated race / ethnicity bucket for a given Housing Advertisement that is negative. This process, known as clamping in academic and industry literature, is

[52] The parameters of DP are defined in the VRS Compliance Metrics Agreement dated January 6, 2023.

implemented to ensure that the count of Ad Impressions after the addition of noise from DP is consistent with the definition of Ad Impressions in the Settlement Agreement. The Settlement Agreement defines Ad Impressions as the "display of ads on Meta platforms," the count of which cannot be negative. Further, academic and industry literature recommend the use of clamping in applications of DP.[53]

In instances where the Actual Impression counts for a given Housing Advertisement are close to zero, there is a potential that the distribution of potential additive noise from DP could result in a negative count of Impressions. In such cases, Actual Impression counts are fixed at zero, which creates a positive skew, on average, for noise added through DP. For impacted Housing Advertisements, the imposition of a lower bound for Actual Impression counts may result in a decrease in Variance, which may translate to an increase in Coverage in data sets where the proportion of Housing Advertisements with relatively low Actual Impression counts is high.

The imposition of a lower bound of zero on Actual Impressions in the VRS Compliance Metrics calculation process is consistent with both the definition of Ad Impressions in the Settlement Agreement and with academic and industry literature on the application of DP, and thus is reasonable. As such, this does not impact Guidehouse's verification of compliance with the VRS Compliance Metrics.

d. Variance and Coverage are sensitive to the BISG probability threshold

Conclusion:

While the BISG probability threshold is a methodology decision that Guidehouse observed may have an impact on Variance and Coverage, Meta's choice of 50% as the BISG probability threshold is consistent with academic, industry, and regulatory best practices, and thus is reasonable.

Supporting Analysis:

To assess Meta's implementation of BISG, Guidehouse used BISG with a 50% probability threshold to assign estimated race / ethnicity to the individuals in the synthetic data and compared the resulting output to the averages of outputs from Meta's 30 BISG synthetic data runs.[54] In Table 8 below, the daily average count of individuals in each race / ethnicity bucket

---

[53] Garrido et al. (2023) provide evidence for clamping in post-processing of DP implementation in the industry. Similarly, Kurz (2021) suggests that, if the DP noise generates values outside the boundaries in a distribution, such values must be clamped. These studies can be found at https://petsymposium.org/popets/2023/popets-2023-0045.pdf and https://rss.onlinelibrary.wiley.com/doi/full/10.1111/1740-9713.01528, respectively.

[54] Meta uses a 50% probability threshold in its implementation of BISG, as described in its November 2021 white paper "How Meta is working to assess fairness in relation to race in the U.S. across its products and systems.". White paper is available at https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems/.

from the Meta runs is compared to the daily average count of individuals in each race / ethnicity bucket per Guidehouse's implementation of BISG with a 50% probability threshold.

Table 8: Average Daily Number of Users with a 50% BISG Probability Threshold

Estimated Race / Ethnicity	Meta	Guidehouse	Difference
White	6,093,214.90	6,093,750.33	535.43
Hispanic	1,852,877.60	1,852,686.09	-191.51
African American	1,090,714.20	1,090,464.31	-249.89
Other	858,846.00	858,752.37	-93.63
Unknown	0.00	-0.15	-0.15
Total	9,895,652.70	9,895,652.94	0.24

The total number of synthetic users aggregated after Meta's BISG implementation was close to the number of synthetic users provided by Guidehouse, resulting in a difference of 0.24 users in an average daily population of 9,895,652.70. The largest observed difference, 535.43 users in the White bucket, was less than 0.01% of the daily average count of users identified as White by Guidehouse (535.43 / 6,093,750.33). The largest difference as a proportion was associated with African American users, where the difference between the two sets of daily users (249.89) was approximately 0.02% of the daily average number of users identified as African American by Guidehouse (249.89 / 1,090,714.20). The differences in counts observed in each bucket are attributable to Meta's implementation of BISG with DP.[55] As Meta's and Guidehouse's implementation of BISG with a probability threshold at 50% resulted in differences in counts of users in each estimated race / ethnicity bucket of at most 0.02% once DP was accounted for, Guidehouse concluded that Meta's implementation of BISG and aggregation of Impressions were consistent with Guidehouse's implementation.

[55] Meta's and Guidehouse's implementation of BISG with a 50% probability threshold did not produce statistically significant differences in results.

Academic, industry, and regulatory literature provide that BISG estimations can be implemented at various probability thresholds, and that higher thresholds produce better predictions. [56] [57] However, a higher probability threshold decreases the number of individuals for whom race / ethnicity can be estimated using BISG. Because of this tradeoff between accuracy and identification, multiple probability thresholds can be considered when implementing BISG. The literature provides 50% - 60% as a range that strikes a good balance between accuracy and identification and is widely used as a best-practice in the financial services industry.[58]

To assess the sensitivity of Variance and Coverage to the BISG probability threshold across this probability threshold range, Guidehouse implemented BISG with a 60% probability threshold using the synthetic data and compared the Variance to that resulting from Meta's implementation of BISG using a 50% probability threshold.

When Guidehouse computed Variance for the synthetic data using race / ethnicity estimated by BISG with a 60% probability threshold, Guidehouse observed a decrease in the average Variance as compared to Meta's average computed Variance for Housing Advertisements with both at least 300 Impressions and greater than 1,000 Impressions.[59]

---

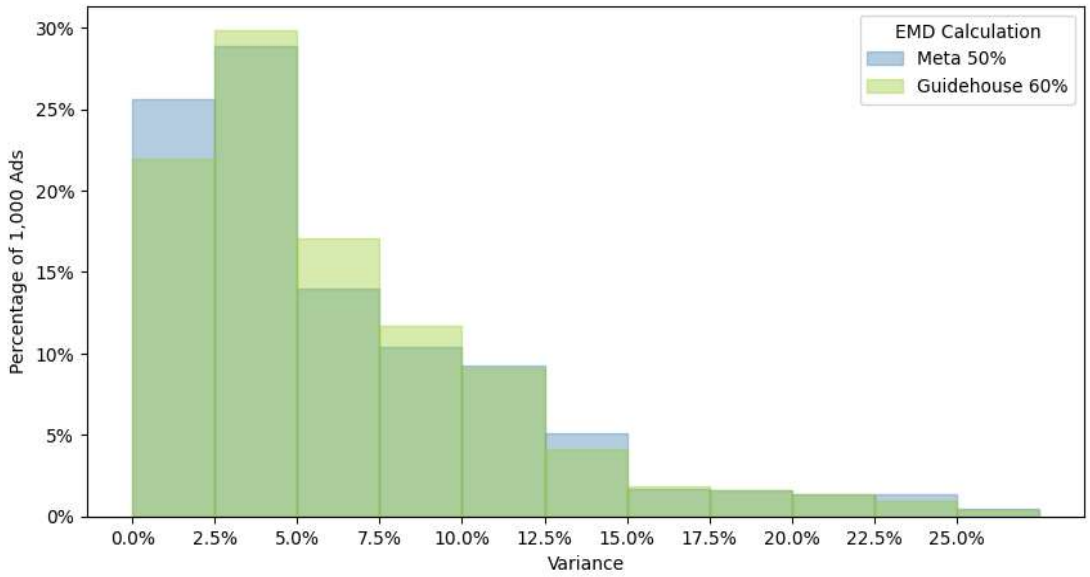
[56] BISG estimation assigns probabilities to each race / ethnicity bucket for a given surname / ZCTA pair. To classify an individual as a single race / ethnicity, a probability threshold is defined. If the probability of an individual being a given race / ethnicity returned by BISG exceeds this probability threshold, the individual is assumed to be that race / ethnicity. There is a tradeoff between the accuracy of the BISG estimation (i.e., a higher probability threshold) and the number of individuals whose race / ethnicity can be assigned by BISG.

[57] Zhang (2018) cites research using a probability threshold no smaller than 50%, but also tests various thresholds and shows that choosing the maximum probability (BISG max) or 80% probability threshold produces more accurate estimates. Paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3169831. Additionally, Chen et al. (2018) shows that choosing the maximum probability over-weights the dominant class ("White" in this sample) in estimation. Jiahao Chen, Nathan Kallus, Xiaojie Mao, Geoffry Svacha, Madeleine Udell, 2018, "Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved" available at https://arxiv.org/pdf/1811.11154.pdf.

[58] CFPB, 2014, "Using publicly available information to proxy for unidentified race and ethnicity" available at https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.

[59] Meta computations include DP, which may also contribute to the disparities.

Figure 5: Comparison of Meta's (50% Probability Threshold and DP) and Guidehouse's (60% Probability Threshold) Variance Distribution



As Figure 5 shows, when a 60% probability threshold is applied to the BISG estimation in the synthetic data, the Variances, on average, change.

This may also translate into an impact to Coverage, as shown in Table 9A and Table 9B.

Table 9A: Comparison of Meta's (50% Probability Threshold with DP) and Guidehouse's (60% Probability Threshold) Variance and Coverage (Housing Advertisements with ≥ 300 Impressions)

	Meta*	Guidehouse	Difference
Average Variance	6.37%	6.31%	0.06%
Coverage at Variance <= 5%	54.06%	52.20%	1.86%
Coverage at Variance <= 10%	78.61%	80.54%	-1.93%

**Average across all of Meta's 30 runs*

Table 9B: Comparison of Meta's (50% Probability Threshold with DP) and Guidehouse's (60% Probability Threshold) Variance and Coverage (Housing Advertisements with > 1,000 Impressions)

	Meta*	Guidehouse	Difference
Average Variance	6.24%	6.09%	0.15%
Coverage at Variance <= 5%	56.80%	56.68%	0.12%
Coverage at Variance <= 10%	78.51%	80.80%	-2.29%

**Average across all of Meta's 30 runs*

In the synthetic data, the average Variance across Housing Advertisements with at least 300 Impressions computed by Guidehouse using a 60% BISG threshold was 6.31% as compared to Meta's computed Variance of 6.37%, creating a 0.06% difference in the mean Variance. When evaluating at the 5% Variance threshold, Guidehouse's computed Coverage was lower than the Coverage computed by Meta, and when evaluating at the 10% Variance threshold, Guidehouse's computed Coverage was higher than the Coverage computed by Meta.

For Housing Advertisements with greater than 1,000 Impressions, the Guidehouse computed Variance was 6.09% as compared to 6.24% when computed by Meta, for a difference of 0.15%. When evaluating at the 5% Variance threshold, Guidehouse's computed Coverage was lower than the Coverage computed by Meta, and when evaluating at the 10% Variance threshold, Guidehouse's computed Coverage was higher than the Coverage computed by Meta.

While this analysis provides that the BISG probability threshold is a methodology decision that may have an impact on Variance and Coverage, Meta's choice of 50% as the BISG probability threshold is consistent with academic, industry, and regulatory best practices, and thus is reasonable.

2. Observations from review of Reporting Period data

a. Meta's Eligible Audience sampling process during the Reporting Period resulted in some Eligible Audience samples that were smaller than the Meta-requested sample size

Conclusion:

Meta leverages an API to sample users into the Eligible Audience for the purposes of calculating the VRS Compliance Metrics.[60] The sampling API is expected to return at least 75% of the requested number of Eligible Users for at least 99.9% of samples selected, which was exceeded in the Reporting Period.[61]

In instances where under sampling did occur, Guidehouse assessed the impact of the under sampling on Coverage for cases that resulted in samples containing 385 users or less for at

---

[60] Meta's sampling module has a target sample size of 6,000 users for Eligible Audiences that include 6,000 or more users or 100% of the estimated Eligible Audience in cases where the Eligible Audience contains fewer than 6,000 users. Meta's sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. The sampling module is designed to raise an alert if more than 1% of the requests to the sampling module within the prior 24 hours delivered a sampled audience of less than 75% of the requested sample size, i.e., an audience of less than 4,500 where the requested size is 6,000 or of less than 75% of the estimated Eligible Audience if it contains fewer than 6,000 total users.

[61] To assess the maximum potential failure rate of the Eligible Audience Sampling API, Guidehouse quantified the total number of Eligible Audience samples by summing the total number of days each Housing Advertisement in the relevant period was active and for each Housing Advertisement where the Eligible Audience sample contained less than 75% of the requested sample size, assumed that all daily samples had fewer than 75% of the requested users.

least one of the daily samples selected for a given Housing Advertisement during the Reporting Period. The analysis showed that Meta exceeded the VRS Compliance Metrics in the Reporting Period by more than the computed potential overestimation of Coverage for both the 5% and 10% Variance thresholds for sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Impressions and Housing Advertisements with more than 1,000 Impressions. As the margin by which Meta exceeded the VRS Compliance Metrics in the Reporting Period is larger than the calculated potential impact of the under sampling, the Eligible Audience sampling observation does not impact Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics for the Reporting Period.

Supporting Analysis:

To understand the impact of the samples in the Reporting Period that did not meet the minimum sample size requested, Guidehouse analyzed the Reporting Period data that covered the period between May 1, 2024 and August 31, 2024. In the Reporting Period, approximately 99.93% of the Housing Advertisements had samples that returned at least 75% of the requested users on all days, and approximately 99.51% of the Housing Advertisements had samples with more than 385 users on all days, which are sufficiently large sample sizes to represent the users in the Eligible Audience.[62] [63]

To assess the maximum potential impact of this under sampling on Coverage, Guidehouse manually adjusted the Variance for Housing Advertisements that had both a sample fulfillment rate of less than 75% of the requested number of users in the sample and / or a sample size less than or equal to 385 users on at least one day in the Reporting Period, which accounted for approximately 0.01% of the Housing Advertisements in the Reporting Period. The manual adjustment ensured that these Housing Advertisements had Variance exceeding both the 5% and 10% Variance thresholds for both sex and estimated race / ethnicity. Once the manual adjustments to Variance for the impacted Housing Advertisements were made, Guidehouse recomputed Coverage for the Reporting Period.

Table 10A shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage computed by Guidehouse after making the manual adjustments as described above (C) for Housing Advertisements with at least 300 Impressions. While manual adjustment of

---

[62] This conclusion is based on a 95% confidence level with a 5% margin of error, which are industry-standard sampling parameters. The confidence level is the probability that the true value being studied falls within a specified range of values. The margin of error denotes the sampling error due to measurement in a sample. 385 is the minimum sample size required for populations as large as 500 million users, which exceeds Meta's daily average usage in the United States. See https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22 for sample size calculations and https://www.sec.gov/ixviewer/ix.html?doc=/Archives/edgar/data/0001326801/000132680124000012/meta-20231231.htm for Meta Annual Report's disclosure of average daily platform usage.

[63] Meta's sampling API is expected to return at least 75% of the requested number of Eligible Users for at least 99.9% of samples selected, which was exceeded in the Reporting Period.

Variance for the under sampled Housing Advertisements resulted in a decrease in Coverage for estimated race / ethnicity at the 10% Variance threshold, the recomputed Coverage still exceeds the VRS Compliance Metrics.

Table 10A: Potential Overestimation of Coverage for Housing Advertisements with ≥ 300 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for Under Sampled Ads (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	90.2%	93.7%	93.7%	Yes
	≤5%	78.3%	86.3%	86.3%	Yes
Estimated Race / Ethnicity	≤10%	80.1%	83.8%	83.7%	Yes
	≤5%	56.8%	64.0%	64.0%	Yes

Table 10B shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage computed by Guidehouse after making the manual adjustments as described above (C) for Housing Advertisements with more than 1,000 Impressions. The manual adjustment of Variance for the under sampled Housing Advertisements did not cause a change in the Coverage across either sex or estimated race / ethnicity.

Table 10B: Potential Overestimation of Coverage for Housing Advertisements with > 1,000 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for Under Sampled Ads (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	91.7%	95.1%	95.1%	Yes
	≤5%	84.5%	89.0%	89.0%	Yes
Estimated Race / Ethnicity	≤10%	81.0%	85.5%	85.5%	Yes
	≤5%	61.0%	68.8%	68.8%	Yes

As Table 10A and Table 10B exhibit, Meta exceeded the VRS Compliance Metrics in the Reporting Period after adjusting for the maximum potential impact of under sampling for both the 5% and 10% Variance thresholds for sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Impressions and Housing Advertisements with more than 1,000 Impressions. As the margin by which Meta exceeded the VRS Compliance Metrics in the

Reporting Period is larger than the calculated potential impact of the under sampling on Coverage, the Eligible Audience samples smaller than the minimum requested sample size do not impact Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics for the Reporting Period.

b. Meta's decisions related to the treatment of unknown ZCTAs, ZCTAs with low populations, Housing Advertisements with small daily Audiences, and unknown sex may result in a subset of Ad Impressions not being captured in VRS Compliance Metrics calculations

Conclusion:

For approximately 0.43% of the Housing Advertisements in the Reporting Period data, there is a larger than 20% absolute difference in the sum of Potential Impressions across sex and estimated race / ethnicity. Similarly, for approximately 1.67% of the Housing Advertisements in the Reporting Period data, there is a larger than 20% absolute difference in the sum of Actual Impressions across sex and estimated race / ethnicity. These observed differences can be attributed to four decisions, enumerated below, which appear reasonable and have a combined impact on the Reporting Period data that was not large enough to impact Coverage.

Supporting Analysis:

As explained by Meta, the discrepancies in the sum of Ad Impressions can be attributed to one or more of the following factors:

1. When a user's ZCTA cannot be estimated by Meta, their race / ethnicity is not estimated using BISG. Rather, they are assigned to the "Unknown" estimated race / ethnicity bucket.
2. When a Housing Advertisement is delivered to a user with a ZCTA that does not have a total population of at least 100 people who are 18 or older, their race / ethnicity is not estimated using BISG. Rather, they are assigned to the "Unknown" estimated race / ethnicity bucket.
3. When a Housing Advertisement has an Eligible Audience or Actual Audience containing fewer than ten unique users in total, Meta does not run BISG on that subset of Ad Impressions, and the user race / ethnicity is not estimated. Rather, they are assigned to the "Unknown" estimated race / ethnicity bucket.
4. When a user does not self-report a sex of either male or female, his / her sex is considered "Unknown."

Any Ad Impressions delivered to users with "Unknown" estimated race / ethnicity are not counted in the VRS Compliance Metrics calculations for estimated race / ethnicity; however, they may be counted in the VRS Compliance Metrics calculations for sex. The converse is true in cases where sex is not known, but race / ethnicity is able to be estimated for an Ad Impression. Ad Impressions omitted for one of the reasons above could potentially impact Variance and Coverage.

Approximately 0.43% of the Housing Advertisements in the Reporting Period had Potential Impression counts that deviated between sex and estimated race / ethnicity by more than 20%. Similarly, approximately 1.67% of the Housing Advertisements in the Reporting Period had Actual Impression counts that deviated between sex and estimated race / ethnicity by more than 20%.

To assess the maximum potential impact of this issue in Impression counts on Coverage, Guidehouse manually adjusted the Variance for Housing Advertisements such that Housing Advertisements that had a deviation of more than 20% in either Potential or Actual Impressions counts between sex and estimated race / ethnicity had Variance exceeding both the 5% and 10% Variance thresholds for both sex and estimated race / ethnicity. Once the manual adjustments to Variance for the impacted Housing Advertisements were made, Guidehouse recomputed Coverage for the Reporting Period.

Table 11A shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage computed by Guidehouse after making the manual adjustments as described above (C) for Housing Advertisements with at least 300 Impressions. While manual adjustment of Variance for the impacted Housing Advertisements resulted in a decrease in Coverage, the margin by which Meta exceeded the VRS Compliance Metrics in the Reporting Period is larger than the calculated potential impact of the Impression inconsistency on Coverage.

Table 11.A: Potential Overestimation of Coverage for Housing Advertisements with ≥ 300 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for Ads with Impression Inconsistency (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	90.2%	93.7%	93.5%	Yes
	≤5%	78.3%	86.3%	86.2%	Yes
Estimated Race / Ethnicity	≤10%	80.1%	83.8%	83.0%	Yes
	≤5%	56.8%	64.0%	63.7%	Yes

Table 11B shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage computed by Guidehouse after making the manual adjustments as described above (C) for Housing Advertisements with more than 1,000 Impressions. While manual adjustment of Variance for the impacted Housing Advertisements resulted in a decrease in Coverage, the margin by which Meta exceeded the VRS Compliance Metrics in the Reporting Period is larger than the calculated potential impact of the Impression inconsistency on Coverage.

Table 11.B: Potential Overestimation of Coverage for Housing Advertisements with > 1,000 Impressions

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for Ads with Impression Inconsistency (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	91.7%	95.1%	95.0%	Yes
	≤5%	84.5%	89.0%	89.0%	Yes
Estimated Race / Ethnicity	≤10%	81.0%	85.5%	85.3%	Yes
	≤5%	61.0%	68.8%	68.7%	Yes

Given the results in Table 11A and 11B above, the issue does not impact Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics for the Reporting Period.

c. Aggregate impact of observations from review of Reporting Period data

Guidehouse assessed the aggregate maximum potential impact of the two observations discussed in Section III.2 – Observations from review of Reporting Period data above on Coverage in the Reporting Period.[64] Guidehouse manually adjusted the Variance for Housing Advertisements that had at least one sample in the Reporting Period with fewer than 75% of the requested sample and / or 385 or fewer Eligible Users and for Housing Advertisements with a 20% or greater difference between either Potential or Actual Impressions for sex and estimated race / ethnicity, such that Variance exceeded both the 5% and 10% Variance thresholds for both sex and estimated race / ethnicity.

Table 12A shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage recomputed by Guidehouse as described above (C) for Housing Advertisements with at least 300 Impressions.

[64] Observations include both impacts of system functionality in the VRS Compliance Metrics calculation process and impacts of certain users not providing sex data or Meta not having sufficient information to estimate a user's ZCTA.

Table 12A: Coverage for Housing Advertisements with ≥ 300 Impressions after Adjusting for the Two Reporting Period Data Observations

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for All Three Observations (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	90.2%	93.7%	93.5%	Yes
	≤5%	78.3%	86.3%	86.1%	Yes
Estimated Race / Ethnicity	≤10%	80.1%	83.8%	83.0%	Yes
	≤5%	56.8%	64.0%	63.7%	Yes

Table 12B shows the target VRS Compliance Metrics (A), Coverage Reported by Meta (B), and Coverage recomputed by Guidehouse as described above (C) for Housing Advertisements with greater than 1,000 Impressions.

Table 12B: Potential Overestimation of Coverage for Housing Advertisements with > 1,000 Impressions after Adjusting for All Two Reporting Period Data Observations

	Variance Threshold	VRS Compliance Metrics (A)	Meta – Reported Coverage (B)	Guidehouse – Calculated Coverage Adjusting for All Three Observations (C)	Meet the Target VRS Compliance Metrics (Is C > A?)
Sex	≤10%	91.7%	95.1%	95.0%	Yes
	≤5%	84.5%	89.0%	89.0%	Yes
Estimated Race / Ethnicity	≤10%	81.0%	85.5%	85.3%	Yes
	≤5%	61.0%	68.8%	68.7%	Yes

While recomputed Coverage is lower than Meta-reported Coverage at both the 5% and 10% Variance thresholds for both sex and estimated race / ethnicity for both Housing Advertisements with at least 300 Impressions and Housing Advertisements with more than 1,000 Impressions, the recomputed Coverage exceeded the VRS Compliance Metric in all cases, and thus these observations do not impact Guidehouse's verification of the VRS Compliance Metrics in the Reporting Period.

IV. Background - Settlement Agreement and Scope of Work

1. Settlement Agreement

On June 27, 2022, Meta entered into a settlement with DOJ.[65] DOJ filed the Settlement Agreement concurrently with a Complaint (Complaint) against Meta alleging violations of the Fair Housing Act (FHA) based on Meta's provision of Housing Advertisement targeting options on the basis of sex and race / ethnicity and the placement of those Housing Advertisements. Meta denied liability and any and all wrongdoing related to these allegations.[66] DOJ designed the Settlement Agreement provisions to resolve the Complaint.

Pursuant to the Settlement Agreement, Meta will:

1. Maintain publishing of active Housing Advertisements in the Ads Library, as required by the March 29, 2019 Settlement Agreement and Release (NFHA Settlement) between Meta and the National Fair Housing Alliance (NFHA), and take reasonable steps to notify users of Meta Platforms that active Housing Advertisements are available to search and view through the Ads Library, pursuant to Settlement Agreement ¶7;

2. Maintain Housing Advertisement identification processes established in the NFHA Settlement and, on the VRS Implementation date and every four months thereafter, submit a report to DOJ and the Reviewer with the number of Housing Advertisements sampled and the number of false positive and false negative Housing Advertisements identified in the reporting period, pursuant to Settlement Agreement ¶8;

3. Maintain limited Housing Advertisement targeting options made available to advertisers, pursuant to the NFHA Settlement. Any new targeting options added to the Housing Ad Flows in accordance with the standards set forth in Settlement Agreement ¶9.a must be shared DOJ, who will have thirty (30) days to review and notify Meta of any objections based on the standards set forth in Settlement Agreement ¶9.a prior to the option being added to Housing Ad Flows, pursuant to Settlement Agreement ¶9.b;

4. Stop delivery of Housing Advertisements targeted using the Special Ad Audience tool by December 31, 2022 and eliminate access to the Special Ad Audience tool and Lookalike Audience tool in Housing Ad Flows, pursuant to Settlement Agreement ¶9.c;

---

[65] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement.

[66] Pursuant to Settlement Agreement ¶5, the Extended Term of the Settlement Agreement will be four (4) years from the Effective Date of the Settlement Agreement. The term of the Settlement Agreement will be the Extended Term, ending on June 27, 2026. The Extended Term is defined in the Joint Letter filed by DOJ on behalf of both DOJ and Meta on January 9, 2023, Dkt. 12.

5. Develop a system, referred to as the VRS, to reduce the Variances in Ad Impressions between the Eligible Audience and Actual Audience for sex and estimated race / ethnicity, pursuant to Settlement Agreement ¶10;

6. Maintain the practice of requiring certification of compliance with anti-discrimination policies and applicable laws for all persons placing Housing Advertisements on Meta Platforms, pursuant to Settlement Agreement ¶11;

7. Maintain the practice of providing enhanced educational content on anti-discrimination policies and applicable laws to all persons placing Housing Advertisements on Meta Platforms, pursuant to Settlement Agreement ¶12;

8. Provide training on FHA to select Meta teams, pursuant to Settlement Agreement ¶13;

9. Make a statement on the Meta website about the Settlement Agreement, its obligations under the Settlement Agreement, and the importance of taking steps to prevent unlawful discrimination on internet platforms, pursuant to Settlement Agreement ¶14; and,

10. Prepare a Compliance Report every four (4) months during the term of the Settlement Agreement verifying compliance with the VRS Compliance Metrics, which will be shared with a third-party Reviewer, pursuant to Settlement Agreement ¶16.

2. Meta's VRS Compliance Metrics

The VRS Compliance Metrics are a measure of the effectiveness of VRS to reduce the Variances in Ad Impressions between the Eligible Audience and the Actual Audience for sex and estimated race / ethnicity, pursuant to Settlement Agreement ¶10, where:

1. Sex will be determined by information reported by users in their Meta profiles; [67]
2. Estimated race / ethnicity will be determined using privacy-enhanced BISG; [68] [69] and,
3. Each user in the Eligible Audience will be weighted by the total number of impressions for any Advertisements displayed to the user on Meta Platforms in the prior thirty (30) days when measuring the Variance between Eligible and Actual Audiences. [70] [71]

The VRS performance is measured using Earth Mover's Distance (EMD), also known as the Wasserstein Metric, and compliance will be determined based on VRS Compliance Metrics.

The VRS Compliance Metrics Agreement date January 6, 2023 defines the "metrics for how much the VRS will reduce any Variances in Ad Impressions between Eligible Audiences and Actual Audiences for sex and estimated race / ethnicity" required by the Settlement Agreement ¶10(b). [72] On January 9, 2023, DOJ and Meta jointly filed a letter with the court advising that they had agreed to the VRS Compliance Metrics and setting forth those agreed-upon metrics. The court then adopted the parties' joint letter as an order. More specifically, VRS Compliance Metrics were set forth as shown in Table 13 and Table 14 below.[73]

---

[67] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement ¶10.a.v.

[68] Meta's BISG implementation process includes adaptations designed to preserve user privacy and prevent the creation of durable records of user race / ethnicity, including obfuscating race / ethnicity buckets during BISG estimation and the addition of DP, or randomized noise, to the data to prevent reidentification of individual data from aggregate data. Meta's application of privacy enhancement is discussed further in white papers available at https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems and https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf.

[69] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement ¶10.a.v.

[70] Ibid., ¶10.a.iv.

[71] Reference to weighting of Eligible Audience users by total impressions of all Advertisements displayed to the user on Meta platforms has been updated to accurately reflect Settlement Agreement ¶10.a.iv This update applies retroactively to all prior publications of the VRS Compliance Metrics Verification Report.

[72] Ibid., ¶10.b.

[73] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7.

Table 13: VRS Compliance Metrics for Housing Advertisements with at least 300 Ad Impressions Delivered in the Reporting Period

	Variance	Coverage		
		By April 30, 2023	By August 31, 2023	By December 31, 2023
Sex	≤10%	80.6%	84.8%	90.2%
	≤5%	68.5%	73.4%	78.3%
Estimated Race / Ethnicity	≤10%	69.7%	74.0%	80.1%
	≤5%	48.5%	52.6%	56.8%

Table 14: VRS Compliance Metrics for Housing Advertisements with more than 1,000 Ad Impressions Delivered in the Reporting Period

	Variance	Coverage		
		By April 30, 2023	By August 31, 2023	By December 31, 2023
Sex	≤10%	82.6%	87.2%	91.7%
	≤5%	73.2%	79.1%	84.5%
Estimated Race / Ethnicity	≤10%	72.2%	76.1%	81.0%
	≤5%	54.3%	57.5%	61.0%

.

From December 31, 2023 through the end of the Extended Term of the Settlement Agreement, Meta agreed to reach the target Coverage ratios set forth under the December 31, 2023 columns in Table 13 and Table 14 above.

Per the VRS Compliance Metrics Agreement dated January 6, 2023, for the three reporting periods in 2023, Meta agreed to include in the VRS Compliance Metrics Housing

Advertisements that both begin and end delivery of Ad Impressions during the given four-month reporting period. For reporting periods beginning in 2024, Meta will include in the VRS Compliance Metrics Housing Advertisements that have ended delivery of Ad Impressions during the given four-month reporting period, regardless of the impression delivery start date.

3. Reviewer's Role and Scope

Guidehouse was proposed by Meta and has the consent of DOJ to serve as the independent third-party Reviewer, pursuant to ¶18 of the Settlement Agreement. The Reviewer is an independent third-party and pursuant to Settlement Agreement ¶17 will "review each Compliance Report and verify compliance with the VRS Compliance Metrics."[74]

For the Reporting Period, Guidehouse verified compliance with the VRS Compliance Metrics by:

1. Assessing the following components of the Meta VRS Compliance Metrics calculation process for accuracy and robustness, using synthetic data created by Guidehouse:[75]
    a. Sampling of the Eligible Audience;
    b. BISG implementation; and,
    c. Aggregation of Eligible Audience and Actual Audience Impressions and the subsequent computation of Variance through EMD and Coverage; and,
2. Confirming that the Variance and Coverage metric calculations for sex and estimated race / ethnicity performed by Meta are accurate, using actual aggregated data provided by Meta to Guidehouse for the Reporting Period.

[74] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement ¶17.

[75] Disaggregated impression data for the Reporting Period is not available so Guidehouse used synthetic data for evaluation of processes requiring individual user- or Impression-level data.

V. Verification Methodology

Guidehouse adopted a two-step verification approach, where the first step assessed components of the VRS Compliance Metrics calculation process using synthetic data, and the second verified the Meta-reported Coverage by independently replicating the calculation steps using aggregated Impression data for Housing Advertisements subject to the VRS Compliance Metrics in the Reporting Period.

1. Step 1: Assessment of VRS Compliance Metrics Calculation Process

Guidehouse assessed the following components of the VRS Compliance Metrics calculation process:

1. Meta's sampling of Eligible Audience for use in calculating the VRS Compliance Metrics;
2. Meta's implementation of BISG to estimate race / ethnicity; and,
3. Meta's aggregation of Potential Impressions and Actual Impressions and the subsequent computation of Variance and Coverage.

To assess these processes, Guidehouse generated a synthetic dataset that contained 10,000,000 last name and ZCTA combinations to identify synthetic users and assigned User IDs and sex to these users. These synthetic users were associated with 10,000 synthetic Housing Advertisements in the dataset.[76] To be able to compute Variance and Coverage, Guidehouse generated Eligible Impression and Actual Impression counts for each synthetic user and synthetic Housing Advertisement in the dataset. Further details about the synthetic dataset generation are presented in Appendix B.

To assess the sampling of users from the Eligible Audience used to calculate the VRS Compliance Metrics, Meta ingested the set of Ad ID and User ID pairs in the synthetic dataset into their sampling process and returned 30 samples of users associated with each synthetic Housing Advertisement. Guidehouse independently selected 30 random samples of users for each Housing Advertisement in the synthetic dataset. Guidehouse computed Variance and Coverage separately for (1) the full synthetic dataset, (2) Meta's samples, and (3) Guidehouse's samples for sex and estimated race / ethnicity and performed comparisons of Variance and Coverage calculated for the synthetic datasets.

Meta and Guidehouse also used the full synthetic dataset to estimate the race / ethnicity of the synthetic users with BISG. Guidehouse then compared aggregated results of BISG estimation to validate Meta's implementation of BISG was consistent with Guidehouse's implementation of

[76] Guidehouse can evaluate components of the VRS Compliance Metrics calculation process using a synthetic dataset with any data distribution. As a starting point, Guidehouse relied on publicly available data from a survey of Meta users and target distributions of Variance for sex and estimated race / ethnicity drawn from the data reported by Meta in conjunction with the Compliance Report issued on May 30, 2024.

BISG and Meta's aggregation of Impressions was consistent with Guidehouse's aggregation. Finally, Guidehouse compared Meta's and Guidehouse's calculation of Variance and Coverage.

2. Step 2: Verification of VRS Compliance Metrics for the Reporting Period

Guidehouse used data compiled by Meta for the Reporting Period to compute Variance and Coverage. Guidehouse then compared the calculated Coverage to the VRS Compliance Metrics for the Reporting Period and the Coverage reported by Meta. Meta provided the data for the Reporting Period in the schema in Figure 6 below.

Figure 6: Meta VRS Compliance Metrics Reporting Schema

						Inputs to Calculate Variance													
						Potential Impressions						Actual Impressions							
				Impression Bucket		Sex		Estimated Race/Ethnicity				Sex		Estimated Race/Ethnicity					
#	Hashed Ad ID	Ad Start Date	Ad End Date	300-1000	>1000	Male	Female	White	Hispanic	African American	Other	Male	Female	White	Hispanic	African American	Other	Variance (Sex)	Variance (Estimated Race / Ethnicity)
1																			
2																			
3																			
...																			
n																			

To compute Variance, Guidehouse calculated the proportion of Potential Impressions and Actual Impressions in Meta's data for each sex and estimated race / ethnicity bucket for a given Housing Advertisement, where the buckets for sex are "Male" and "Female" and for estimated race / ethnicity are "White," "Hispanic," "African American," and "Other," pursuant to the VRS Compliance Metrics Agreement dated January 6, 2023.[77] To calculate the proportion, Guidehouse took the Potential Impression count and Actual Impression count in each sex and estimated race / ethnicity bucket for a given Housing Advertisement and divided them by the total Potential Impression count and total Actual Impression count for that Housing Advertisement, respectively. For example, if there are 600 and 400 Potential Impressions for male and female, the ratios would be 60% (600/1,000) and 40% (400/1,000), respectively.

Using these ratios, Guidehouse summed the absolute differences in ratios between Potential and Actual Impressions separately for sex and estimated race / ethnicity, and divided this sum by two to calculate Variance:

Variance (Sex) = $(|Ratio_{p,m} - Ratio_{e,m}| + |Ratio_{p,f} - Ratio_{e,f}|) \div 2$, and

Variance (Estimated Race / ethnicity) = $(|Ratio_{p,w} - Ratio_{e,w}| + |Ratio_{p,h} - Ratio_{e,h}| + |Ratio_{p,a} - Ratio_{e,a}| + |Ratio_{p,o} - Ratio_{e,o}|) \div 2$,

[77] "VRS Compliance Metrics Agreement." 6 Jan. 2023.

where p and e denote "Potential Impressions" and "Actual Impressions," m and f denote "male" and "female," and $w, h, a,$ and o denote "White," "Hispanic," "African American," and "Other," respectively.

Finally, Guidehouse computed the Coverage by finding the percentage of Housing Advertisements with calculated Variance below the 5% and 10% Variance thresholds defined in the VRS Compliance Metrics Agreement dated January 6, 2023.[78]

[78] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 12.

Appendix A – Definitions

The capitalized terms listed below will have the following meaning, consistent with their definitions in the Settlement Agreement ¶3, 9, 10, 16, and 17 and the January 6, 2023 VRS Compliance Metrics Agreement, unless otherwise noted: [79] [80]

Actual Audience: All users in an Eligible Audience to whom at least one Impression of a Housing Advertisement is displayed.

Ad Impressions or Impressions: Display of ads on Meta Platforms, or any potential or synthetic ads not displayed on Meta Platforms.[81]

Ads Library: An interface that allows users to search and view active Housing Advertisements by advertiser or by location targeting options selected by advertisers.

Compliance Report: Meta-prepared report confirming that it has met the VRS Compliance Metrics for the previous four-month reporting period.

Coverage: The percentage of Housing Advertisements where the Variance is less than or equal to the prescribed Variance threshold.

Differential Privacy: A privacy-enhancing technology that protects against re-identification of individuals within aggregated datasets by adding randomized noise.[82]

Effective Date: The Effective Date of the Settlement Agreement, or the date upon which the Settlement Agreement is entered by the Court or an application to enter the Settlement Agreement is granted, whichever occurs first, as recorded on the Court's docket.

---

[79] United States v. Meta Platforms, Inc. f/k/a Facebook, Inc., 22 Civ. 5187 (JGK), Dkt. No. 7, Settlement Agreement ¶3, 9, 10, 16, 17.

[80] "VRS Compliance Metrics Agreement" 6 Jan. 2023.

[81] Definition of term expanded beyond that of the Settlement Agreement for the purposes of discussing Potential Impressions not displayed to Meta Platforms' users or synthetic Impressions in Guidehouse-generated synthetic data.

[82] Meta's discussion of Differential Privacy is available in white papers https://ai.facebook.com/research/publications/how-meta-is-working-to-assess-fairness-in-relation-to-race-in-the-us-across-its-products-and-systems and https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf.

Eligible Audience: All users who (1) fit targeting options selected by an advertiser for an ad, and (2) received one or more Impressions of any type of ad on Meta Platforms during the last thirty days.

FHA-Protected Classes: Race, color, religion, sex, disability, familial status, and national origin within the meaning of the FHA.

Housing Ad Flows: Interfaces that advertisers use to create Housing Advertisements for publication on Meta Platforms.

Housing Advertisement: An advertisement offering a specific opportunity to rent, lease, sell, hold, convey, transfer, or buy a residential dwelling, and / or offering a specific real-estate related transaction such as residential mortgage, homeowner's insurance, or home appraisal services within the meaning of FHA.

Lookalike Tool: Legacy tool available to advertisers on Meta platforms to create audiences, now replaced by the Special Ad Audience tool.

Meta Platforms: Facebook, Instagram, and Messenger.

Reviewer: An independent third-party responsible for reviewing each Compliance Report and verifying compliance with the VRS Compliance Metrics.

Special Ad Audience: A tool in Housing Ad Flows that allows advertisers to create audiences with commonalities to a group of users, such as the advertisers' current customer, visitors to their websites, or people who like their Facebook page.

Variance: The distance between the potential Impression distribution for the Housing Advertisement and the actual Impression distribution for the Housing Advertisement, for both sex (Male, Female) and estimated race / ethnicity (White, Hispanic, African American, and Other) separately, measured using Earth Mover's Distance.

Variance Reduction System (VRS): A Meta-developed system designed to reduce the Variance in Ad Impressions between Eligible Audiences and Actual Audiences for sex and estimated race / ethnicity.

VRS Compliance Metrics: Metrics agreed upon by DOJ and Meta and filed with the Court on how much the VRS will reduce any Variances in Ad Impressions between Eligible Audiences and Actual Audiences for sex and estimated race / ethnicity.

Guidehouse

Appendix B – Synthetic Data Creation

For the purposes of assessing Meta's selection of a sample of users from the Eligible Audience, its implementation of BISG, its aggregation of Impressions, and calculation of Variance and Coverage, Guidehouse created a synthetic dataset, comprised of 10,000,000 synthetic users and matched those synthetic users with 10,000 synthetic Housing Advertisements.

To create the synthetic dataset representing the 10,000,000 users, Guidehouse performed the following steps:

1. Leveraged Census files to tabulate distribution of population counts of each race / ethnicity for the top 10,000 surnames across the U.S. and percent of each race / ethnicity within ZCTAs.
2. Built dictionaries to store target proportions for sex, estimated race / ethnicity (White, Hispanic, African American, Other), and usage (Frequent, Casual, Infrequent) based on publicly available demographic survey data estimating Meta's user base. [83]
3. Built dictionaries to store weighted likelihood of selecting a surname or ZCTA based on race / ethnicity (e.g., likelihood to be white given a certain surname, and likelihood to be white given a certain ZCTA).
4. Generated a list of 10,000,000 unique User IDs.
5. Randomly assigned (with replacement) surnames for each User ID by sampling from distributions derived from 2010 U.S. Census data for surname frequency by race / ethnicity. This sampling was weighted based on target demographic proportions for race / ethnicity (55% White, 20% Hispanic, 15% African American, and 5% Other), with specific ZCTA and surname assignments weighted by the dictionaries described in Step 3. These targets were derived from publicly available demographic survey data used to approximate Meta's user base.[84]
6. Randomly assigned each User ID a sex based on target demographic proportions (54% Female, 46% Male). These targets were derived from publicly available demographic survey data used to approximate Meta's user base.[85]
7. Assigned each synthetic User ID a ZCTA. ZCTAs were weighted by population within each ZCTA, leveraging 2010 U.S. Census population data. Only included eligible ZCTAs (non- P.O. box ZCTAs, non-territories).

---

[83] Guidehouse leveraged publicly available survey data from a survey of Meta users to develop target parameters as a starting point for synthetic data distribution.

[84] Ibid.

[85] Ibid.

8. Categorized each User ID as having Frequent, Casual, or Infrequent usage of Meta platforms based on target proportions for each race / ethnicity stored in the dictionary as described in Step 1. Target proportions were derived from publicly available demographic survey data on platform usage by race.[86]
    a. Guidehouse leveraged publicly available usage data for U.S. Facebook users across race / ethnicity to model the likelihood a user may be categorized as a Frequent, Casual, or Infrequent user across each estimated race / ethnicity.[87]
    b. Guidehouse did not incorporate sex in categorizing on usage as Guidehouse found no readily available public sources of data with usage by sex.
9. Based on this usage categorization, randomly assigned the number of synthetic Housing Advertisements a user is eligible for based on three separate uniform distributions.
    a. Infrequent users are eligible to see anywhere from 1 to 15 Housing Advertisements.
    b. Casual users are eligible to see anywhere from 16 to 49 Housing Advertisements.
    c. Frequent users are eligible to see anywhere from 50 to 99 Housing Advertisements.

To assign synthetic users to the 10,000 synthetic Housing Advertisements, Guidehouse took the following steps:

1. For each of the 10,000 unique Ad IDs in the synthetic dataset, leveraged data reported by Meta in conjunction with their Compliance Report dated May 30, 2024 to assign targets for Eligible Impressions, Actual Impressions, and Variance range for both sex and estimated race / ethnicity.
2. For each of the synthetic Housing Advertisements, randomly assigned an Eligible Audience size of between approximately 10,000 and 90,000 users.
3. Assigned synthetic users to each synthetic Housing Advertisement. For each synthetic user-synthetic Housing Advertisement pair, assigned Eligible and Actual Impressions counts to achieve the targets assigned in Step 1 above. Guidehouse achieved this through an iterative process, which resulted in some of the 10,000,000 synthetic users not being assigned to any Housing Advertisement in the synthetic dataset.

---

[86] Ibid.

[87] The publicly available Meta user data was limited to Facebook users only. The target proportions are assumed to reflect all synthetic users across the three platforms (Facebook, Messenger, and Instagram).

4. After building the aggregate dataset with a unique record for each synthetic user-synthetic Housing Advertisement pair, disaggregated the synthetic data into ten datasets representing ten separate days of Advertisement delivery.

Figure B1 and Figure B2 below depict the distribution of Variance in the synthetic data created through the steps described above for the Reporting Period, overlayed on the distribution of Variance in Meta's Reporting Period data associated with the May 30, 2024 Compliance Report.

Figure B1: Distribution of Variance for Estimated Race / Ethnicity for Housing Advertisements in Meta's Fourth Reporting Period Data Associated with the May 30, 2024 Compliance Report and Guidehouse's Synthetic Data

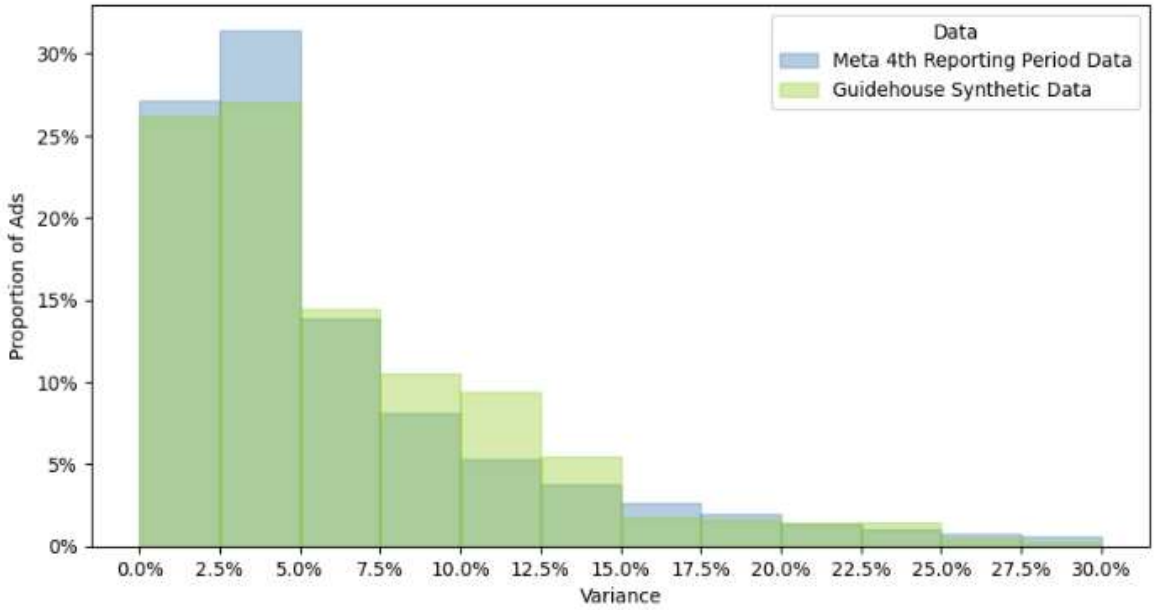
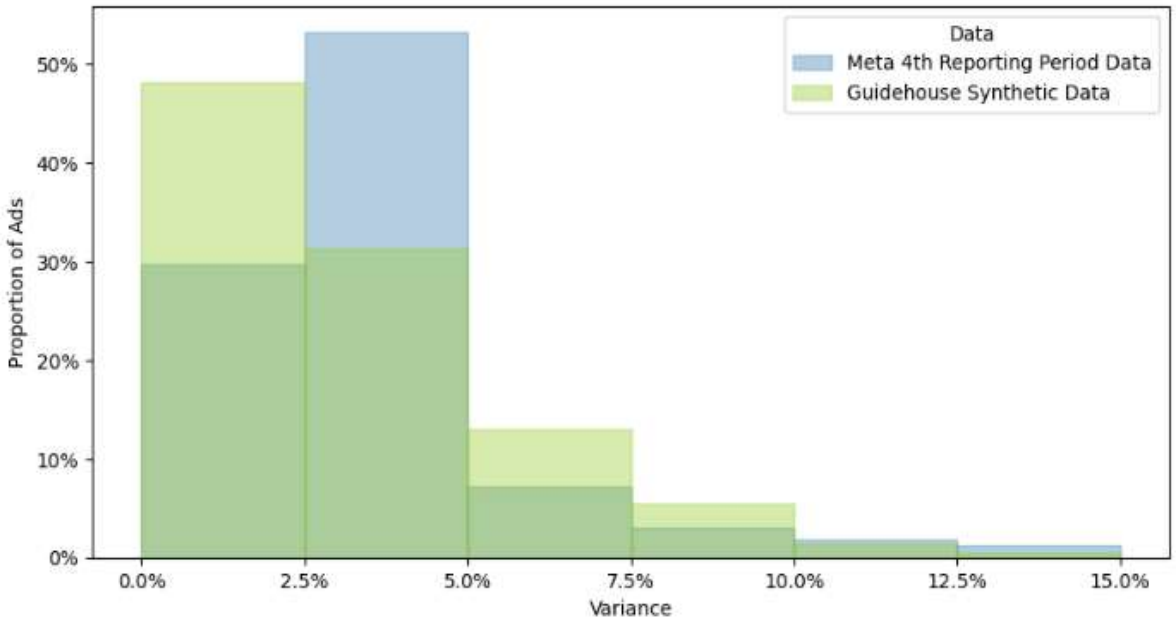


Figure B2: Distribution of Variance for Sex for Housing Advertisements Meta's Fourth Reporting Period Data Associated with the May 30, 2024 Compliance Report and Guidehouse's Synthetic Data



The synthetic data creation steps produced the synthetic dataset that Guidehouse used in the structure in Table B1 below:

Table B1. Synthetic Dataset View for the Fifth Reporting Period

Synthetic Ad ID	Day	Synthetic User ID	Sex	Total Actual Impressions for the User Across All Advertisements	Surname	ZCTA	Number of Actual Impressions (for the Specified Housing Advertisement)
1	1	100	M	8,694	M####Y	21044	0
10	1	10000	F	6,690	B####T	91941	1
10000	2	1000000	M	1,624	S####H	79938	2