Guidehouse
Outwit *Complexity*

∞ Meta

VRS Compliance Metrics Verification
October 30, 2023

Table of Contents

I.   Executive Summary

Guidehouse Inc. (Guidehouse or Reviewer) was proposed by Meta Platforms, Inc. (Meta) and had 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, Guidehouse reviewed the Meta Compliance Report dated September 29, 2023 for the reporting period from May 1, 2023 to August 31, 2023 (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 Meta's 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. 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 in the U.S. across its products and systems".[4] [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 found here: 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 such calculations to the VRS Compliance Metrics established in the VRS Compliance Metrics Agreement dated January 6, 2023 and 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.

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

	Variance Threshold	VRS Compliance Metrics	Meta – Reported Coverage[9]	Guidehouse – Calculated Coverage[10]	Difference in Coverage
Sex	≤10%	84.8%	93.5%	93.5%	0.0%
	≤5%	73.4%	86.7%	86.7%	0.0%
Estimated Race / Ethnicity	≤10%	74.0%	80.1%	80.1%	0.0%
	≤5%	52.6%	57.6%	57.6%	0.0%

---

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

[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.

[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, 2023.

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

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

	Variance Threshold	VRS Compliance Metrics	Meta – Reported Coverage[11]	Guidehouse – Calculated Coverage[12]	Difference in Coverage
Sex	≤10%	87.2%	94.5%	94.5%	0.0%
	≤5%	79.1%	89.1%	89.1%	0.0%
Estimated Race / Ethnicity	≤10%	76.1%	81.0%	81.0%	0.0%
	≤5%	57.5%	61.6%	61.6%	0.0%

Notwithstanding the verification of Meta's compliance with the VRS Compliance Metrics, Guidehouse had five observations as a result of 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.

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

Second, Guidehouse found Variance and Coverage calculated with the synthetic data to be sensitive to the sample size of Eligible Audience members used when calculating the VRS Compliance Metrics. Specifically, as the sample size increased, Coverage computed in the samples converged to the Coverage computed using the full synthetic dataset. The VRS Compliance Metrics Agreement dated January 6, 2023 specifies that the Eligible Audience will be measured based on a sample of users and Meta's minimum sample size threshold of 3,750 users is sufficient for populations as large as 500 million users, which exceeds the average

---

[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, 2023.

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

number of daily Meta platform users and thus is a reasonably sized sample of users.[13] [14] As such, Guidehouse's verification of Meta's compliance with the VRS Compliance Metrics in the Reporting Period is not impacted by this observation.

Third, based on Guidehouse's analysis of the synthetic data, the noise added from DP impacted Meta's calculation of Variance and Coverage for the synthetic data. Meta explained that the effect of the DP noise, which is implemented as a privacy protecting measure, on calculated Variance is inversely related to the difference between the Potential Impression distribution and Actual Impression distribution. Meta also provided empirical evidence that DP noise increased the Variance on average, and due to the distribution of Variance observed in the Reporting Period data, DP is not expected to result in an increase in Coverage, and thus does not impact Guidehouse's verification of the VRS Compliance Metrics.

Fourth, 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.

Guidehouse's fifth observation is based on its analysis of the Reporting Period data as it pertained to 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 ZIP Codes or sex, ZIP Codes 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 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. Please see https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22

[14] Meta disclosed in its 2021 Annual Report that the average daily active users in the U.S. and Canada were approximately 195 million between December 2019 and December 2021. Please see https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/2023/2021-Annual-Report.pdf

II. Verification of VRS Compliance Metrics

For the Reporting Period, Guidehouse 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, in accordance with the Settlement Agreement and the VRS Compliance Metrics Agreement.

In Table 3 and Table 4 below, Guidehouse has 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.[15] The difference in Coverage 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	Meta – Reported Coverage[16]	Guidehouse – Calculated Coverage[17]	Difference in Coverage
Sex	≤10%	84.8%	93.5%	93.5%	0.0%
	≤5%	73.4%	86.7%	86.7%	0.0%
Estimated Race / Ethnicity	≤10%	74.0%	80.1%	80.1%	0.0%
	≤5%	52.6%	57.6%	57.6%	0.0%

---

[15] Compliance Report pursuant to *United States v. Meta Platforms, Inc.*, No. 22-Civ-5187 (S.D.N.Y.) for May 1 - August 31, 2023.

[16] 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, 2023.

[17] 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	Meta – Reported Coverage[18]	Guidehouse – Calculated Coverage[19]	Difference in Coverage
Sex	≤10%	87.2%	94.5%	94.5%	0.0%
	≤5%	79.1%	89.1%	89.1%	0.0%
Estimated Race / Ethnicity	≤10%	76.1%	81.0%	81.0%	0.0%
	≤5%	57.5%	61.6%	61.6%	0.0%

---

[18] 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, 2023.

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

III.   Observations

Through its verification of the VRS Compliance Metrics for the Reporting Period, Guidehouse made five observations, four based on its analysis of the synthetic data and one 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 yields a distribution sex and estimated race / ethnicity across sampled users that is consistent with random sampling. Therefore, Guidehouse confirms 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.[20] [21] For each Housing Advertisement in the synthetic data, Meta selected 30 samples of between 3,888 and 5,000 users each using the same sampling process deployed in the VRS Compliance Metrics calculation process.[22] Similarly, Guidehouse performed a random sampling of exactly 5,000 users for each Housing Advertisement within the synthetic dataset 30 times.

---

[20] 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 fit 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.

[21] 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.

[22] Meta uses a maximum sample size of 5,000 in its computation of VRS Compliance Metrics. Its sampling module has a distributed architecture, where the module divides the data into multiple partitions each handled by an independent task. When a request is made to the module, it aggregates responses from each task to construct the result. During this aggregation process, it is possible that not all tasks respond back within the specified time limit. Therefore, the sample size can be less than 5,000 in certain instances.

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 proportion of synthetic users that are female are 54.43%, 54.41%, and 54.42% of the Eligible Audience population, Meta's samples, and Guidehouse's samples, respectively.

Figure 1. Breakdown of Synthetic Users by Sex
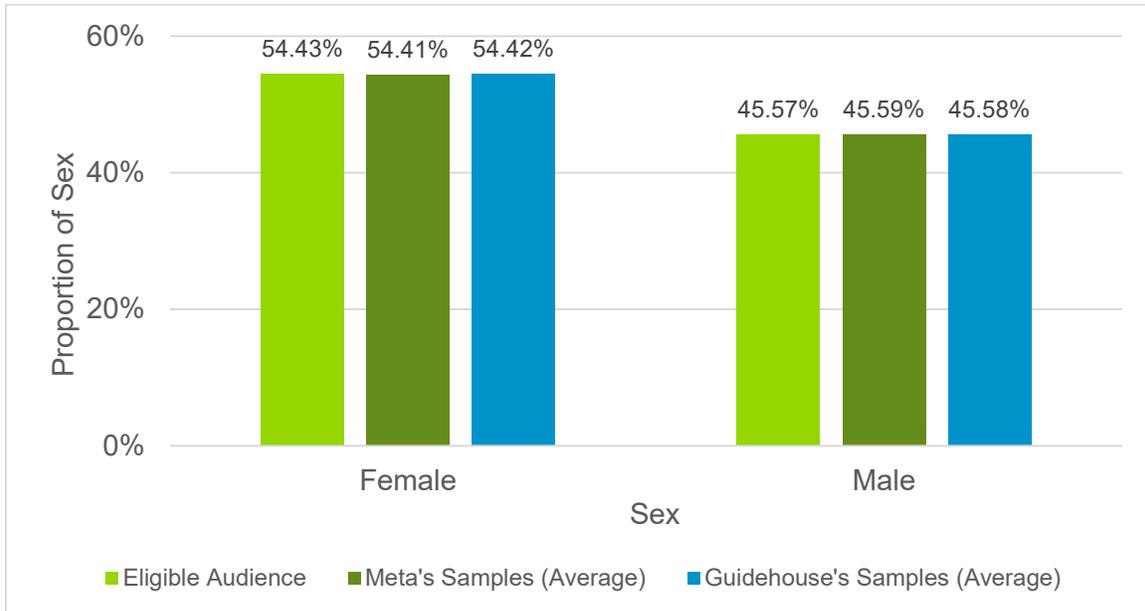
Similarly, Figure 2 below demonstrates the breakdown of synthetic users by estimated race / ethnicity in the synthetic Eligible Audience data, Meta's samples, and Guidehouse's samples. The percentage associated with each estimated race / ethnicity is consistent in the synthetic Eligible Audience dataset and in both samples. For example, the proportion of synthetic users that are categorized as "White" are 57.36%, 57.42%, and 57.36% of the Eligible Audience population, Meta's samples, and Guidehouse's samples, respectively.

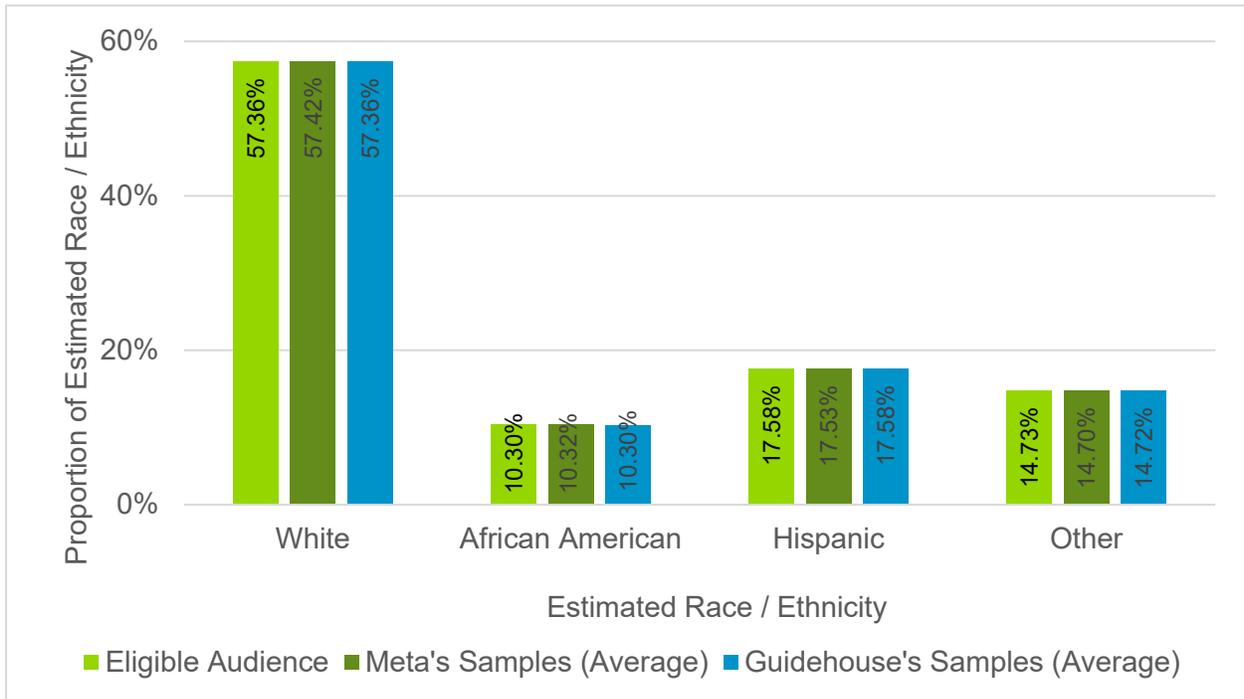**Figure 2. Breakdown of Synthetic Users by Estimated Race / Ethnicity**

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

b. *Size of the Sample of Eligible Audience members may impact Variance and Coverage*

<u>Conclusion:</u>

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. The magnitude of this impact decreases as the size of the sample converges to the Eligible Audience. While Variance and Coverage are measures of Impressions, the VRS Compliance Metrics Agreement dated January 6, 2023 specifies that the Eligible Audience will be measured based on a sample of users and Meta's minimum sample size threshold of 3,750

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.[23] [24] [25]

<u>**Supporting Analysis:**</u>

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.

Table 5 below demonstrates that, at the 5% Variance threshold, the Coverage for sex in the Eligible Audience, Meta's samples, and Guidehouse's samples were 100%, 99.9%, and 99.9%, respectively. Coverage for estimated race / ethnicity in the Eligible Audience, Meta's samples, and Guidehouse's samples were 39.2%, 43.6%, and 41.4%, respectively.

Table 5: 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

	Sex	Estimated Race / Ethnicity
Eligible Audience	100.0%	39.2%
Meta's Samples*	99.9%	43.6%
Guidehouse's Sample*	99.9%	41.4%

Average across 30 samples generated by Meta and Guidehouse separately

As Variance for sex in the synthetic data was clustered around 1%, Guidehouse did not observe an impact in Coverage at the 5% Variance threshold.

For estimated race / ethnicity, Guidehouse tested separately whether Meta's average Coverage for estimated race / ethnicity (43.6%) and Guidehouse's average Coverage for estimated race / ethnicity (41.4%) across the 30 samples was statistically different than the Eligible Audience Coverage for estimated race / ethnicity, and found that average Coverage across both Meta's and Guidehouse's 30 Eligible Audience samples was statistically different from the Coverage

---

[23] 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. Please see https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=500000000&x=98&y=22

[24] 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 and declare itself "unavailable" if less than 75% of tasks respond in time. This prevents the sampling module from delivering a sampled audience of less than 3,750 where the requested size is 5,000.

[25] Meta disclosed in its 2021 Annual Report that the average daily active users in the U.S. and Canada were approximately 195 million between December 2019 and December 2021. Please see https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/2023/2021-Annual-Report.pdf

observed in the full synthetic Eligible Audience.[26] This observation provides evidence that the Coverage estimated in a sample of the Eligible Audience may differ from the Coverage calculated for the full Eligible Audience.

Table 6 exhibits consistent impact on Coverage calculated at the 10% Variance threshold.

Table 6: 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

	Sex	Estimated Race / Ethnicity
Eligible Audience	100.0%	74.5%
Meta's Samples*	100.0%	78.6%
Guidehouse's Sample*	100.0%	76.2%

Average across 30 samples generated by Meta and Guidehouse separately

To understand why average Coverage for estimated race / ethnicity was higher in both Meta's and Guidehouse's samples than the Coverage calculated for the full synthetic Eligible Audience, Guidehouse analyzed the distribution of estimated race / ethnicity of users as compared to the distribution of estimated race / ethnicity for Ad Impressions in the full synthetic Eligible Audience, Meta-selected samples, and Guidehouse-selected samples.

Guidehouse observed that because Ad Impressions were not distributed uniformly across users in the synthetic dataset, the distribution of Impressions across estimated race / ethnicity may vary from the distribution of users across estimated race / ethnicity.

To assess the impact of the difference in distribution between users and Impressions in samples of varying sizes on Variance and Coverage, Guidehouse estimated Variance and Coverage in two sample sizes alternative to the 5,000 used in Meta's VRS Compliance Metrics reporting process and compared the Variance and Coverage to those computed in the synthetic Eligible Audience data. The alternative sample sizes assessed assumed a fixed proportion of 30% and 50% applied to all synthetic Housing Advertisements in the synthetic Eligible Audience dataset. Since the number of users eligible for an Advertisement in the synthetic data varied between 28,875 and 89,220, the sample sizes in the 30% and 50% samples varied accordingly between 8,662 and 26,766 and 14,437 and 44,610 synthetic users, respectively, for a synthetic Housing Advertisement.[27]

---

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

[27] 28,875 * 30% = 8,662.5 and 89,220 * 30% = 26,766.
28,875 * 50% = 14,437.5 and 89,220 * 50% = 44,610.

Table 7 below shows the estimated Variance in the Eligible Audience, Meta's samples, and Guidehouse's samples for sizes of 5,000, 30%, and 50% of the full Eligible Audience.

Table 7: Comparison of Variance in Guidehouse's Samples of Various Sample Sizes

	Sex	Estimated Race / Ethnicity
Eligible Audience (BISG Threshold: 50%)	0.89%	6.07%
Meta (Sample Size: ~5K, BISG Threshold: 50%)[28]	1.10%	6.09%
Guidehouse (Sample Size: 5K, BISG Threshold: 50%)	1.10%	6.10%
Guidehouse (Sample Size: 30%, BISG Threshold: 50%)	0.93%	6.09%
Guidehouse (Sample Size: 50%, BISG Threshold: 50%)	0.90%	6.07%

The sample size changes have marginal impact on Variance for both sex and estimated race / ethnicity, with Variance for both sex and estimated race / ethnicity converging to the Eligible Audience population mean as the sample size increases.

Guidehouse also analyzed the impact of the sample size on Coverage in Table 8 below. As Coverage is a measure of the portion of a population with Variance below a given threshold, it can be sensitive to changes in Variance, particularly when Variance in the population is clustered around the defined threshold. Since the Variance estimates for sex are clustered around 1%, changing the sample size did not create a large enough change in Variance to impact the Coverage estimates at the 5% or 10% Variance thresholds. While the average Variance for estimated race / ethnicity did not change significantly when the sample size increased, the average Coverage across 30 samples decreased from 41.4% (in 30 samples of 5,000) to 39.7% (in 30 samples, where sample size is 50% of the Eligible Audience for each synthetic Housing Advertisement), converging to the population Coverage.

Table 8: Comparison of Coverage in Guidehouse's Samples of Various Sample Sizes

	Sex	Estimated Race / Ethnicity
Eligible Audience (BISG Threshold: 50%)	100.0%	39.2%
Meta (Sample Size: ~5K, BISG Threshold: 50%)[29]	99.9%	43.6%
Guidehouse (Sample Size: 5K, BISG Threshold: 50%)	99.9%	41.4%
Guidehouse (Sample Size: 30%, BISG Threshold: 50%)	100.0%	40.3%
Guidehouse (Sample Size: 50%, BISG Threshold: 50%)	100.0%	39.7%

---

[28] Meta sample size varied between 3,888 users and 5,000 users per sample, with an average of 4,475 users per sample observed in the synthetic data analysis.

[29] Ibid.

These findings provide empirical evidence that the difference in Coverage between the full Eligible Audience and the Eligible Audience sample averages decrease as the number of users sampled increases.

c. DP adds noise that may impact Variance and Coverage

Conclusion:

DP noise added by Meta in its implementation of BISG may potentially impact Variance and Coverage; however, the behavior of that noise is not expected to increase Coverage in the Reporting Period data, as the concentration of Housing Advertisements immediately below the 5% and 10% Variance thresholds is higher than the concentration immediately above the thresholds.[30]

Supporting Analysis:

To evaluate the impact of DP on Variance and Coverage, Guidehouse generated synthetic user and Housing Advertisement data 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.[31] 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 Guidehouse's and Meta's Variance distributions.

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.

---

[30] User race / ethnicity is not self-reported information in the Meta user database. As such, 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 .

[31] Meta's and Guidehouse's use of a 50% BISG probability threshold, aggregation of the data, and computation of Variance and Coverage were the same in this analysis to isolate the impact of DP.

Figure 3. Comparison of Meta's (with DP) and Guidehouse's (without DP) Variance Distribution
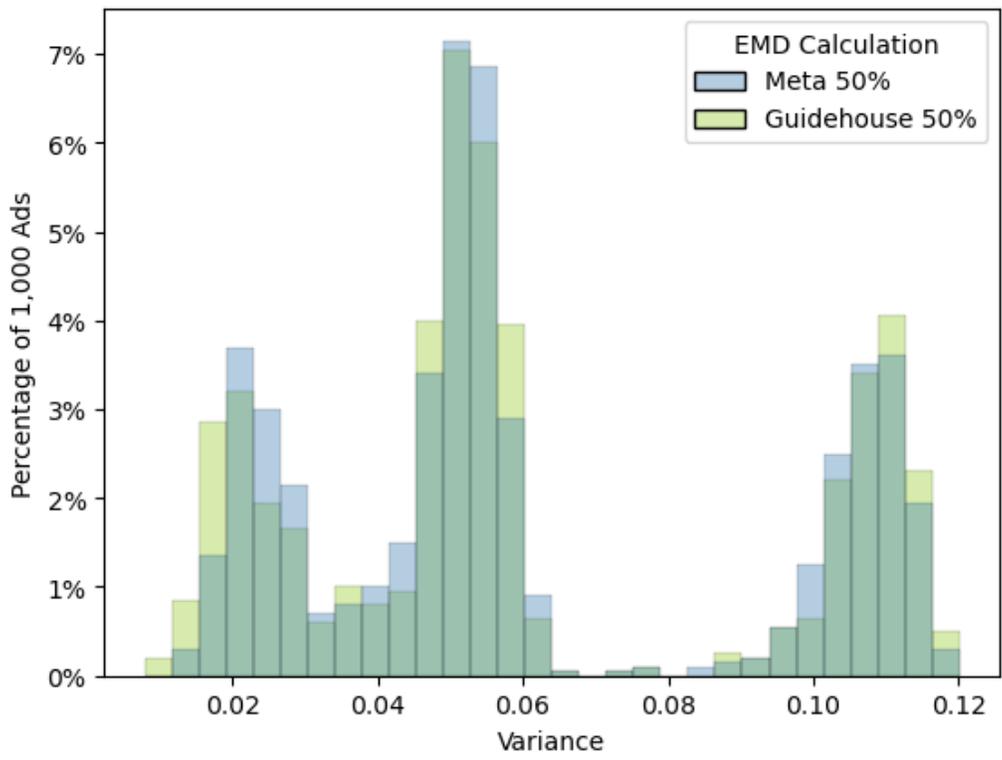
The average Variance computed by Meta across all advertisements in the synthetic data was 6.2%, versus an average Variance of 6.1% computed by Guidehouse. The minimum and maximum average Variance calculated by Meta was 0.9% and 12.3%, respectively, as compared to 0.8% and 12.0% computed by Guidehouse.

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 4.74% and a maximum computed Variance of 4.87%, or a spread of 0.13%. The Housing Advertisement with the largest observed fluctuation in Variance had a minimum computed Variance of 7.35% and a maximum computed Variance of 9.38%, or a spread of 2.03%. While the magnitude of the potential impact of DP on Variance decreased when compared to the results from the First Reporting Period, similarly, 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 9. As discussed above, the average Variance in the synthetic data calculated by Guidehouse was 6.1%, with approximately 44% of Housing Advertisements with Variance within one percentage point of the 5% Variance threshold used to determine Coverage. Consequently,

a very small amount of noise applied to the Variance for these Housing Advertisements could move them from one side of the 5% Variance threshold to the other, resulting in a large impact on Coverage at the 5% Variance threshold.

Table 9: Comparison of Meta's (with DP) and Guidehouse's (without DP) Variance and Coverage

	Meta*	Guidehouse	Difference
Variance	6.2%	6.1%	0.1%
Coverage at Variance <= 5%	36.6%	39.2%	-2.6%
Coverage at Variance <= 10%	73.7%	74.5%	-0.8%

Average across all of Meta's 30 runs

As Table 9 shows, Guidehouse's computed Coverage with the synthetic data at the 5% Variance threshold was 39.2%, compared to Meta's average Coverage of 36.6% across 30 runs.[32] Therefore, at the 5% Variance threshold, the difference in Variance caused a difference of -2.6% in Coverage between Guidehouse's calculation and Meta's average calculation. At the 10% Variance threshold, Guidehouse's computed Coverage was 74.5%, compared to Meta's average Coverage of 73.7% across 30 runs, resulting in a difference of -0.8% in Coverage between Guidehouse's calculation and Meta's average calculation.

Based on these results, Guidehouse observed that DP may potentially 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.

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.[33] 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

---

[32] Meta's Coverage across 30 runs ranged between 36.1% and 37.2% at the 5% Variance threshold.

[33] Meta's analysis consisted of first adding DP noise to Potential Impression distributions and Actual Impression distributions for 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 true value of Variance for each Advertisement. Meta then added DP noise one additional time to the assumed true value for each Advertisement and calculated the average difference in Variance between the second application of DP and the assumed true value for the 100 runs.

potential distribution of Variance with DP may be less than the Variance without DP. If true Variance 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 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.[34]

Figure 4: Distribution of Variance for Housing Advertisements in Meta's Reporting Period Data
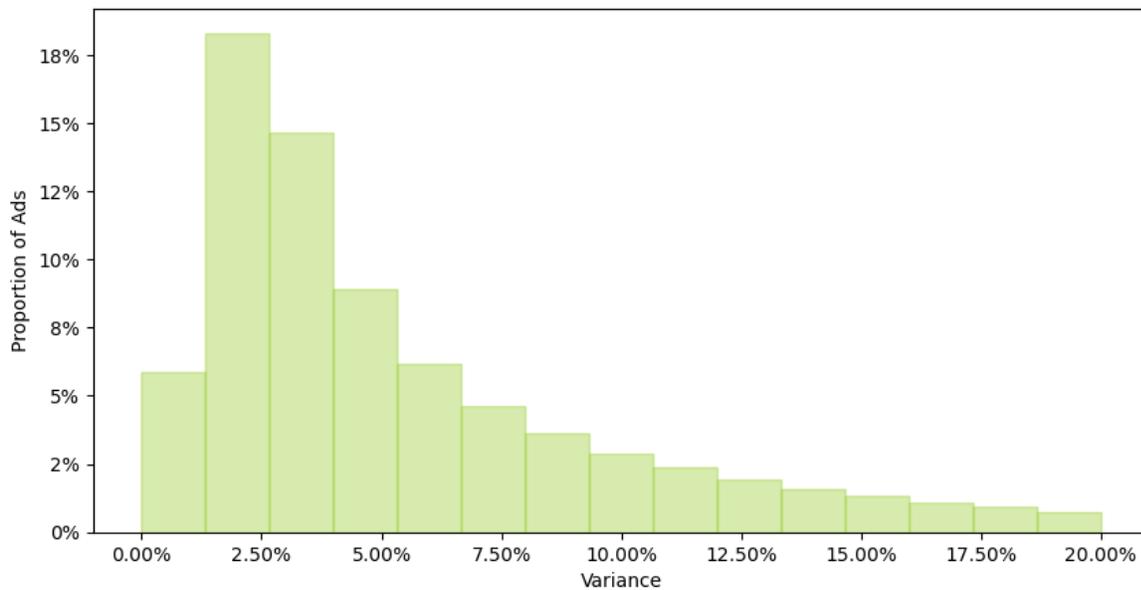
As shown in Figure 4, Variance observed in Meta's Reporting Period data is clustered below both the 5% and 10% Variance thresholds. Further, for both the 5% and 10% Variance thresholds, the number of Housing Advertisements with observed Variance immediately below the threshold outweighed the number of 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.

---

[34] Figure 4 displays the distribution of all housing ads within the reporting period with a Variance below 20%, truncating outliers to show the distribution more clearly for ads near the Coverage thresholds.

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

Conclusion:

While the BISG probability threshold is a methodology decision that Guidehouse has 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.[35] In Table 10 below, the average count of individuals in each race / ethnicity bucket from the Meta runs is compared to the count of individuals in each race / ethnicity bucket per Guidehouse's implementation of BISG with a 50% probability threshold.

Table 10: Comparison of Synthetic Data Output of BISG with a 50% Probability Threshold

Estimated Race / Ethnicity	Meta	Guidehouse	Difference
White	561,564.9	555,468.0	6,096.9
Hispanic	182,868.1	181,904.0	964.1
African American	105,324.0	104,416.0	908.0
Other	149,871.3	157,838.0	-7,966.7
Unknown	374.2	374.0	0.2
Total	1,000,002.5	1,000,000.0	2.5

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 2.5 users in a population of 1,000,000. Guidehouse noted that there were larger differences across estimated race / ethnicity buckets relative to the difference in total population. Meta provided analyses and data to substantiate that, while Meta's and Guidehouse's use of Census data were consistent, the differences in counts in each bucket were attributable to a combination of differences in pre-processing of the Census data and the impact of DP. As the results of Meta's and Guidehouse's implementation of BISG with a probability threshold at 50% were similar once pre-processing and DP were accounted for, Guidehouse concluded that Meta's implementation of BISG and aggregation of Impressions were consistent with Guidehouse's implementation.

---

[35] 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/.

Academic, industry, and regulatory literature provide that BISG estimations can be implemented at various probability thresholds, and that higher thresholds produce better predictions. [36] [37] 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.[38]

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, which relies on race / ethnicity estimated by BISG with a 50% probability threshold.[39]

---

[36] BISG estimation assigns probabilities to each race / ethnicity bucket for a given surname / ZIP Code 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.

[37] 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*.

[38] 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*.

[39] 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
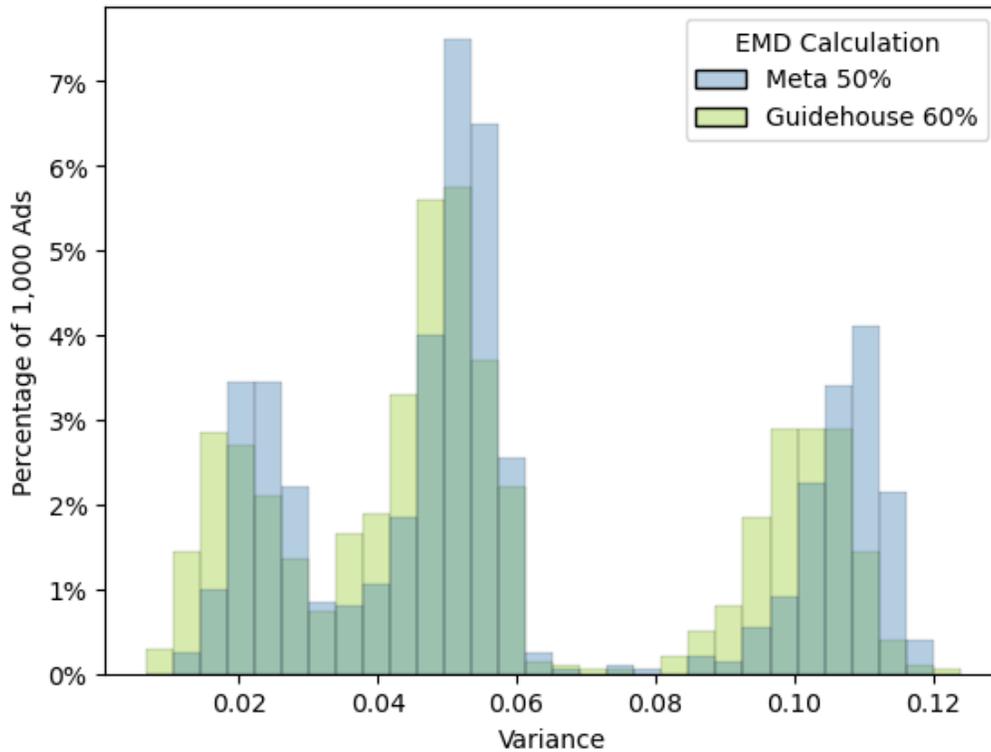
As Figure 5 shows, when a 60% probability threshold is applied to the BISG estimation in the synthetic data, the Variances, on average, decreased. More specifically, Guidehouse's Variance estimates using a 60% BISG probability threshold were, on average, lower than those calculated by Meta using a 50% BISG probability threshold.

This may also translate into an impact to Coverage, as shown in Table 11.

Table 11: Comparison of Meta's (50% Probability Threshold with DP) and Guidehouse's (60% Probability Threshold) Variance and Coverage

	Meta*	Guidehouse	Difference
Average Variance	6.2%	5.7%	0.5%
Coverage at Variance <= 5%	36.6%	48.5%	-11.9%
Coverage at Variance <= 10%	73.7%	81.4%	-7.7%

Average across all of Meta's 30 runs

In the synthetic data, the average Variance across all Housing Advertisements computed by Guidehouse was 5.7% as compared to Meta's computed Variance of 6.2%, creating a 0.5% difference in the mean Variance. When evaluating at both 5% and 10% Variance thresholds, 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 decisions related to the treatment of unknown ZIP Codes, ZIP Codes 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

<u>Conclusion:</u>

For less than 1% 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 less than 1% of the Housing Advertisements in the Reporting Period data, there is a larger than 20% absolute difference in 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.

<u>Supporting Analysis:</u>

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 ZIP Code is not known 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 ZIP Code 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 for a given day, 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.

Less than 1% of Housing Advertisements in the Reporting Period had Ad Impression counts that deviated between sex and estimated race / ethnicity by more than 20% and the majority of Housing Advertisements with deviations greater than 20% had Variance exceeding both the 5% and 10% thresholds. As Meta-reported Coverage met the VRS Compliance Metrics by margins greater than 1%, these Housing Advertisements would not impact Meta's compliance with the VRS Compliance Metrics.

IV. Background - Settlement Agreement and Scope of Work

1. Settlement Agreement

On June 27, 2022, Meta entered into a settlement with DOJ.[40] 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, race, and ethnicity and the placement of those Housing Advertisements. Meta denied liability and any and all wrongdoing related to these allegations.[41] 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;

---

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

[41] 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; [42]

2. Estimated race / ethnicity will be determined using privacy-enhanced BISG; [43] [44] and,

3. Each user in the Eligible Audience will be weighted by the total number of impressions for any Housing Advertisements displayed to the user on Meta Platforms in the prior thirty (30) days when measuring the Variance between Eligible and Actual Audiences. [45]

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 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). [46] 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 12 and Table 13 below.[47]

---

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

[43] Meta's BISG implementation process includes adaptations designed to preserve user privacy and prevent the creation of a 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 discuss 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.

[44] Ibid., ¶10.a.v.

[45] Ibid., ¶10.a.iv.

[46] Ibid., ¶10.b.

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

Table 12: 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 13: 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 12 and Table 13 above.

Per the VRS Compliance Metrics Agreement, 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 intends to 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 had 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."[48]

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:[49]

   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,

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.

---

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

[49] Disaggregated impression data for the Reporting Period is not available, so synthetic data is used 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 the Variance.

To assess these processes, Guidehouse generated a synthetic dataset that contained 1,000,000 last name and ZIP Code combinations to identify synthetic users and assigned User IDs and sex to these users. These synthetic users were associated with 1,000 synthetic Housing Advertisements in the dataset.[50] 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 the 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 the 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 BISG and Meta's aggregation of Impressions was consistent with Guidehouse's aggregation.

---

[50] 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.

Finally, Guidehouse compared the Variance and Coverage estimated separately by Meta and Guidehouse.

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

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

Figure 6: Meta VRS Compliance Metrics Reporting Schema

#	Hashed Ad ID	Ad Start Date	Ad End Date	Impression Bucket		Inputs to Calculate Variance													Variance (Sex)	Variance (Estimated Race / Ethnicity)
						Potential Impressions						Actual Impressions								
						Sex		Estimated Race/Ethnicity				Sex		Estimated Race/Ethnicity						
				300-1000	>1000	Male	Female	White	Hispanic	African American	Other	Male	Female	White	Hispanic	African American	Other			
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 race / ethnicity bucket for a given Housing Advertisement, where the buckets for sex are "Male" and "Female" and for race / ethnicity are "White," "Hispanic," "African American," and "Other," pursuant to the VRS Compliance Metrics Agreement.[51] To calculate the proportion, Guidehouse took the Potential Impression count and Actual Impression count in each sex and 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:

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

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

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.

---

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

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.[52]

---

[52] 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: [53] [54]

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.[55]

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.[56]

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.

---

[53] 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.

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

[55] 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.

[56] Meta's discussion of Differential Privacy is available at privacytech.fb.com/differential-privacy/ and 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.

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.

Appendix B – Synthetic Data Creation

For the purposes of assessing Meta's selection of a sample of Eligible Audience, its implementation of BISG, and its aggregation of Impressions, Guidehouse created a synthetic dataset, comprised of 1,000,000 synthetic users and 1,000 synthetic Housing Advertisements.

To create the synthetic dataset representing the 1,000,000 users, Guidehouse performed the following steps:

1. 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.

2. Generated a list of 1,000,000 unique User IDs.

3. 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. [57] This sampling was weighted based on target demographic proportions for race / ethnicity (55% White, 20% Hispanic, 15% African American, and 5% Other). These targets were derived from publicly available demographic survey data used to approximate Meta's user base.[58]

4. 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.[59]

5. Assigned each synthetic User ID a ZIP Code. ZIP Codes were weighted by population within each ZIP Code, leveraging 2010 U.S. Census population data. Only included eligible ZIP Codes (non- P.O. box ZIP Codes, non-territories).

6. 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 one. Target proportions were derived from publicly available demographic survey data on platform usage by race.[60]

---

[57] Please see https://www.census.gov/data/developers/data-sets/surnames.html

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

[59] Ibid.

[60] Ibid.

   a. Guidehouse leveraged publicly available usage data for U.S. Facebook users across ethnicity to model the likelihood a user may be categorized as a Frequent, Casual, or Infrequent user across each estimated race / ethnicity.[61]

   b. Guidehouse did not incorporate sex in categorizing on usage as Guidehouse found no readily available public sources of data with usage by sex.

7. 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 79 Housing Advertisements.

   c. Frequent users are eligible to see anywhere from 80 to 99 Housing Advertisements.

To create the synthetic dataset representing the 1,000 Housing Advertisements on Meta platforms, Guidehouse took the following steps:

1. Generated a list of 1,000 unique AD IDs.

2. For each of the synthetic Housing Advertisements, randomly assigned an Eligible Audience size of between approximately 30,000 and 90,000 users.

3. For each synthetic Housing Advertisement and User ID pair, assigned the number of Actual Impressions based on the user's usage category (Frequent, Casual, Infrequent) using a random uniform distribution.

4. Once Guidehouse confirmed the synthetic data matched demographic distributions established above, Guidehouse iteratively introduced additive noise to actual impressions for select users to test that Meta's sampling process and implementation of BISG are distribution agnostic and work under a variety of data distributions.

These steps produced the synthetic dataset that Guidehouse used in the structure in Table B1 below:

---

[61] 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, Instagram).

Table B1. Synthetic Dataset View for the Second Reporting Period

Synthetic Ad ID	Synthetic User Id	Sex	Total Actual Impressions for the User Across All Advertisements	Surname	ZIP Code	Number of Actual Impressions (for the Specified Housing Advertisement)
002203CC50B7451A9C168C5B8223621B	8FMT78BCB5NFT72XWQQEAZTU4K8VBBJR	Male	8,694	M####Y	21044	0
002203CC50B7451A9C168C5B8223621B	6F7VSY4S0RWAYF8IHFDROTI43PB6676U	Female	6,690	B####T	91941	18
002203CC50B7451A9C168C5B8223621B	2EKJDN31CY5ZPJGO9MLNYU7NWG8DYJU3	Male	1,624	S####H	79938	2