Memorandum

To: Vicki Sandiford, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency

From: Leland Deck and Megan Lawson, Stratus Consulting Inc.

Date: 2/3/2010

Subject: Statistical analysis of existing urban visibility preference studies

During the CASAC meeting on October 5-6, 2009, Dr. Bill Malm and other CASAC members suggested that a limited dependent variable statistical analysis could be used to analyze the acceptability criteria responses in the four cities for which there are existing urban visibility preference studies. It was the view of those Panel members that successful statistical analyses of the studies results would provide an estimate of a "best fit" central tendency function describing the results of the preference studies, as well as confidence intervals around the estimated functions. Such analyses would also make it possible to conduct hypothesis testing, such as examining whether the estimated 50% criteria level in one study is statistically different than the 50% criteria level in another study.

On the basis of the CASAC comments and the information available in the previous Stratus Report (Stratus Consulting, 2009), EPA concluded it was appropriate to conduct further statistical analyses on the available urban visibility preference studies. Subsequently, EPA asked Stratus Consulting to re-examine the data from these studies and identify several methods for statistical analyses along the lines CASAC members suggested. This memorandum provides a description of the statistical analyses we conducted, and summarizes the results.

Data

While we do not have complete original response data from each preference study, certain data available in all four studies can be used to derive a set of data for an analysis comparing the results from each of the four[1] cities. This available data is the percentage of respondents that rated each individual photograph (or image) as acceptable. We also know the total number of individuals that rated each photograph, as well as the haziness level in each photograph, measured in deciviews (dv). Using these pieces of information we were able to assemble a master data set of 19,280 observations from the original data. Each observation is associated

[1] In the initial set of analyses discussed in this memorandum we combine the results from the 2001 Washington, DC focus group study with all 26 participants in the "Test 1" analysis from Smith and Howell (2009). "Test 1" was designed to replicate the 2001 focus group study, with a goal of making two sets of results directly comparable. Additional analysis described later in this memorandum uses a different set of statistical techniques to examine the Washington, DC studies in more detail.

with an individual binary "yes" or "no" acceptability answer, the dv level, and the city location for a single photograph.

For example, in the Phoenix study 385 participants rated each of 21 different WinHaze images. Hence the Phoenix study contributes 8,085 (385 × 21) observations, nearly 41.9% of the total set of 19,280 observations in the master data set. The 32 photographs used in the Denver study contribute 6,848 observations (35.5% of the total), the 20 photographs in the British Columbia contribute 3,600 observations (18.7% of the total), and the combined Washington, DC studies (combining data from the DC-2001 study with the Test 1 data from the DC-2009 study) contribute 747 (3.9% of the total). The 19,280 observations are fairly evenly split, with 9,452 "yes" observations, and 9,828 "no" responses.

The participants in each study viewed a series of images with different dv levels. While the data collected by the original researchers included information linking each individual with their ratings on each picture, such detailed information is currently only available for the Washington, DC study conducted in 2009. Access to this additional level of information in the 2009 Washington study allows us to conduct an additional type of analysis accounting for individual heterogeneity of preferences regarding acceptable levels of visibility.

Statistical Analysis Models

All of the analyses described in this memorandum are logistic regressions using the logit model. The logit model is a generalized linear model used for binomial regression analysis which fits explanatory data about binary outcomes (in this case, a person rating a photograph acceptable or not) to a logistic function curve.

In the context of the preference studies, the logit model estimates the function that best approximates the percentage of respondents that will rate a photograph acceptable based on a set of explanatory variables. The observations on the dependent variable have one of two discrete values: 1 (the person rated the photograph acceptable) or 0 (unacceptable). In our context, the logit model estimates the proportion of participants who will find any particular dv level acceptable. In our analysis, there were two basic types of explanatory (independent) variables; one continuous numerical variable (the photograph's haziness level in dv), and a set of discrete variables that identify which city the observation is from. We estimate two variations of the logit model, using the basic explanatory variables in different ways.

The fundamental form of a logistic function is:

$$probability("yes") = f(z) = \frac{1}{1+e^{-z}}.$$

where the variable z, known as the logit, is the influence of all the explanatory variables:

$$z = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \varepsilon.$$

In our analysis the estimated logistic function f (z) is the estimated probability of the participants in the study rating a photograph acceptable, given the dv value of the photograph and what city the observation came from.

We conducted the logit analysis using two alternative forms of the logit model.

Model 1 is a simple form of the logit model, and includes the dv value and uses the city information to create a set of categorical indicator variables. This analysis assumes that all respondents have a similar shape to their response function (the probability function of responding "yes" given the dv level of a photograph), but investigates whether the location of the response function differs in the four cities.

The logit for Model 1 is:

$$z = Intercept + \beta_1 dv + \beta_2 BC + \beta_3 DC + \beta_4 Phoenix + \varepsilon .$$

The variables BC (British Columbia), DC (Washington), and Phoenix are the indicator (or "dummy" variables. For example, the BC variable is set equal to one if the observation is from the BC study, and set to zero if that observation is from a study in a different city study. Denver is used as the omitted city indicator variable, allowing the estimated coefficients on the other three city indicator variables to estimate if the response function is different in those cities than in Denver. The term ε represents the error with which the model was estimated, or the difference between the actual and predicted values of *z*. The logit model assumes that ε has a mean of zero.

The Model 1 form of the logit model estimates a single "slope" for the response function in all cities as β_1, the coefficient for haziness (dv). The other terms shift the intercept. The intercept for Denver is simply the estimated parameter *Intercept*. The effective intercept for the other cities becomes the sum of Intercept plus the coefficient on the city's indicator variable, for example the intercept for Washington is *Intercept* + β_3.

Model 1 creates one test of the hypothesis that the responses in each city are the same. If the estimated coefficient on a particular city variable is statistically significant, the analysis would imply that the city's response function is likely shifted relative to the Denver function, and that city would have a different dv value for the 50% criteria. A positive and significant city coefficient shifts that city's response function to the right, resulting in the dv level where 50% criteria level in that particular city is higher than Denver's.

Model 2 is a more general model than Model 1, and relaxes the assumption in Model 1 that the slope of the response function is the same in every city. Model 2 includes not only dv and the city indicator variables as in Model 1, but also a set of interaction terms, where each city dummy variable is multiplied by the dv level. The logit for Model 2 is:

$$z = Intercept + \beta_1 dv + \beta_2 BC + \beta_3 (dv \times BC) + \beta_4 DC$$
$$+ \beta_5 (dv \times DC) + \beta_6 Phoenix + \beta_7 (dv \times Phoenix) + \varepsilon.$$

For example, in Model 2 the estimated total intercept for Washington becomes *Intercept* + β_4, and the estimated slope of the Washington function is $\beta_4 + \beta_5$.

In the fully interacted Model 2 a statistically significant estimate of the city indicator variable coefficients (β_2, β_4, or β_6) has the same implication as in Model 1; the response function is likely shifted relative to the Denver function. A statistically significant estimate of the interaction term coefficient (β_3, β_5, or β_7) for a particular city implies that the response function has a different slope than the Denver function.

The fully interacted model produces the same results as conducting a separate logit analysis for each of the four cities. The interacted model, however, makes it easier to conduct hypothesis testing on the estimated mean response functions.

The predicted mean dv values at each of the acceptance criteria presented here are a function of the coefficients on dv and the other explanatory variables, each of which have their mean and standard deviation. Therefore, a confidence interval constructed around this predicted mean must account for both the variance and covariance of the parameter estimates. Using a Monte Carlo estimation approach, we made 1000 random draws from the joint distribution of the coefficients using the mean vector and variance-covariance matrix of the parameter estimates for the distribution parameters. For each of these draws we then calculated the predicted mean dv. After removing the lower and upper 5% of the simulated values, the lower and upper end of the range of predicted values represent the lower and upper range of the 95% confidence interval. Confidence intervals calculated using this procedure are known as Krinsky-Robb confidence intervals (Krinsky and Robb, 1986). Because estimating Krinsky-Robb confidence intervals requires a separate Monte Carlo analysis for each acceptability criteria dv level, we only estimate confidence intervals for five different acceptability levels: 90%, 75%, 50%, 25%, and 10%.

The Krinsky-Robb procedure assumes that the estimated parameters are normally distributed, which may or may not be true. To explore the potential impact of this assumption, for one logit analysis we also conducted an alternative procedure that does not assume a normal distribution. This alternative procedure (Hole, 2007) uses a bootstrap method to estimate the confidence intervals for the estimated mean 50% criteria. The confidence intervals using the bootstrap were within 1% of the confidence intervals using the Krinsky-Robb procedure, indicating that the multivariate normal assumption imposed by the Krinsky-Robb procedure is not unreasonable. We also conducted hypothesis tests using the median dv values estimated using the bootstrapping procedure. The conclusions from these hypothesis tests were identical to the conclusions from the other hypothesis tests.

Statistical Analysis Results, Inter-City analyses

We conducted all the logit analyses described in this document using STATA® Data Analysis and Statistical Software (Release ES 10.1), using the LOGIT procedure. The Krinsky-Robb analysis used STATA's "wtpcikr" module. The bootstrap method (Hole, 2007) was conducted using STATA's "bootstrap" module.

Model 1 Results, Inter-City Analysis

Table 1 presents the parameter estimates from the logit analysis with city indicators (Model 1) which effectively shift the intercept. The Washington, DC data in this analysis includes both DC-2001 and DC-2009 (Test 1) data. The Denver study is the omitted indicator city in this analysis, so the intercept term coefficient for Denver is equal to the Constant. The intercept for the other cities is the sum of the constant plus the coefficient for the respective city. The coefficient for variable dv is the estimated slope for all four cities.

Table 1. Model 1 logit analysis results

Variable	Coefficient (β)	Standard error	z-statistic	Pr |β| = 0	5% confidence estimate	95% confidence estimate
dv	-0.4187	0.0059	-71.09	< 0.001	-0.430	-0.407
British Columbia	1.1164	0.0630	17.72	< 0.001	0.993	1.240
Washington, DC	3.8743	0.1325	29.25	< 0.001	3.615	4.134
Phoenix	1.8021	0.0576	31.31	< 0.001	1.689	1.915
Constant	8.3073	0.1186	70.07	< 0.001	8.075	8.540

McFadden's pseudo-R^2 for the Model 1 estimate[2] was 0.474.

[2] While pseudo-R^2 is, like traditional R^2, bounded between zero and one, it does not have the same interpretation. R^2 can be interpreted as the percentage of the variation in the dependent variable explained by variation in the independent variables. Pseudo-R^2, on the other hand, is the percent improvement in log likelihood from using the full set of explanatory variables, relative to a model that uses only a constant. It offers a sense for how much better the model fits when the explanatory variables are added, but cannot tell us the percentage of variation we are explaining. Pseudo R^2, instead of traditional R^2, must be used in evaluating logit and other maximum likelihood estimation models. Similar to R^2, a higher pseudo-R^2 indicates a model with a better fit.

The Log likelihood chi^2 test strongly rejects the null hypothesis there is no effect of explanatory variables on the probability that a respondent would find a photograph acceptable (Pr (chi^2) = 0 < 0.000).

The z-statistic (also known as the Wald z-statistic) in a logit analysis is analogous to the t-statistic in a conventional linear regression. The z-statistic is simply the ratio of the estimated coefficient to its standard error, and can be used to estimate the probability that the estimated coefficient is equal to zero. The column in Table 1 labeled "Pr |β| = 0" is the 2-tailed p-value used in testing the null hypothesis that the estimated parameter is zero. The Pr |β| values shown in Table 1 are all less than 0.005 ("~0"), indicating that all of the estimated coefficients are very statistically significant. Because the city dummy variables are significant, in Model 1 we reject the hypothesis that the four studies have an identical response function.

Figure 1 shows the estimated response functions in each city for the logit analysis with city indicators, as well as the underlying data as was shown in Figure 14 of the Stratus Consulting final report (Stratus Consulting, 2009). While Model 1 estimates the shape of a response function that is identical in each city, the positive and significant coefficients on the city variables in Model 1 result in the response functions for the different cities to shift to the right of the Denver function.

The logit analysis results also support estimating the dv value where the 50% acceptability criteria are met in each city. The 50% acceptability criteria occur at the level of haziness where half the survey participants said the visibility is acceptable, and half said it was not acceptable. In Figure 1, the 50% criteria level is the dv value where the estimated response function crosses the 50% response level on the y axis.

As a sensitivity analysis, it is also possible to calculate the dv levels that meet alternative decision criteria. For example, one can calculate the estimated dv level at which 75% of the participants said the visibility was acceptable. This 75% criterion would occur at better visibility (i.e., lower dv values) than the 50% criteria. Similarly, one can also calculate the estimated the dv level that any desired percentage of the participants said was acceptable. The Model 1 estimates of alternative acceptability criteria dv values for each city are shown in Table 2.

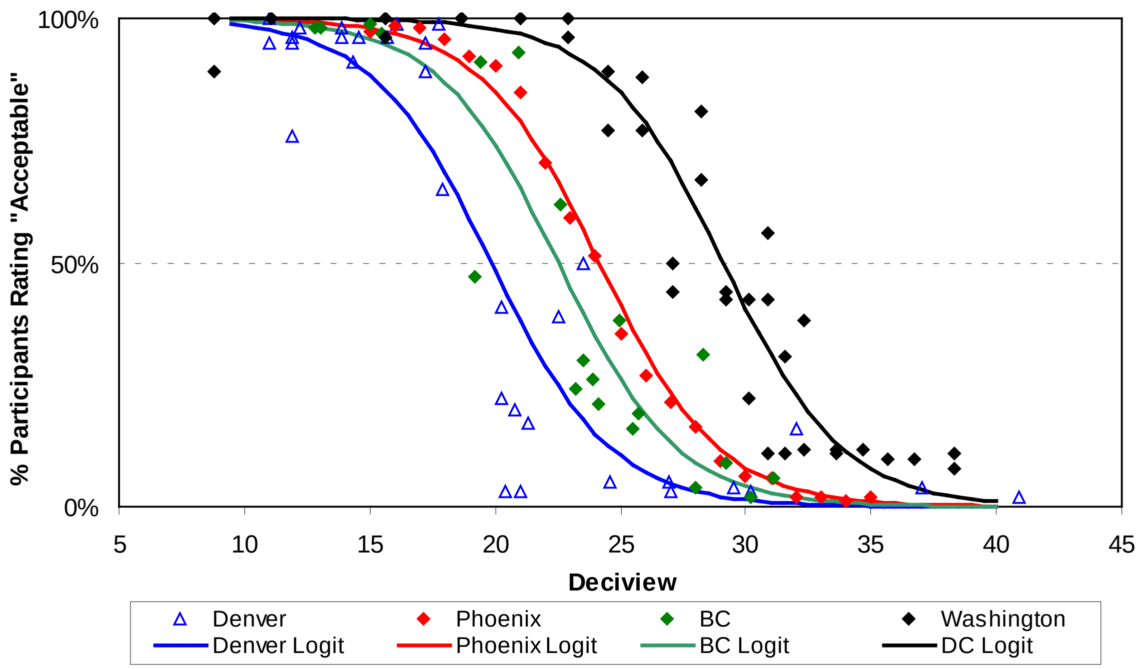

Figure 1. Estimated response functions for full cities using the Model 1 logit analysis.

Table 2. Model 1 estimated haziness (dv) levels of alternative acceptability criteria

	Denver	British Columbia	Washington, DC	Phoenix
90% acceptability criteria	14.59	17.26	23.85	18.90
75% acceptability criteria	17.22	19.88	26.47	21.52
50% acceptability criteria	19.84	22.51	29.10	24.15
25% acceptability criteria	22.47	25.13	31.72	26.77
10% acceptability criteria	25.09	27.76	34.34	29.39

The range of the Model 1 estimates of the 50% acceptability criteria is very consistent with the Candidate Protection Level (CPL) range of 20 dv to 30 dv identified in the U.S. EPA (2009) report *Particulate Matter Urban-Focused Visibility Assessment; External Review Draft* (UFVA).

Model 2 Results, Inter-City Results

Table 3 presents the parameter estimates from the fully interacted logit analysis, which investigates whether both slope and the intercept of the estimated response function differ between cities. Denver was again used as the omitted city in the fully interacted model.

Table 3. Model 2 logit analysis results

Variable	Coefficient (β)	Standard error	z-statistic	Pr |β| = 0	5% confidence estimate	95% confidence estimate
dv	-0.3862	0.0094	-41.16	< 0.001	-0.4045	-0.3678
British Columbia	1.0496	0.3589	2.92	0.003	0.3463	1.7530
Washington, DC	2.9450	0.8458	3.48	< 0.001	1.2873	4.6026
Phoenix	3.5682	0.3015	11.84	< 0.001	2.9773	4.1591
BC × dv	-0.0029	0.0162	-0.18	0.860	-0.0345	0.0288
Wash. × dv	0.0200	0.0293	0.68	0.495	-0.0374	0.0774
Phoenix × dv	-0.0797	0.0136	-5.88	< 0.001	-0.1063	-0.0531
Constant	7.6844	0.1830	41.99	< 0.001	7.3257	8.0431

The pseudo-R^2 for the Model 2 estimate was 0.4756 (very similar to the Model 1 results), and the Model 2 log likelihood chi^2 test also strongly rejects the null hypothesis there is no effect of the explanatory variables on the probability that a respondent would find a photograph acceptable (Pr (chi^2) = 0 < 0.000).

The city indicator coefficients in this full interaction model are all positive and statistically significant, as they were in Model 1, indicating that the response functions for different cities shifted right (relative to Denver). However, of all the interactions only the Phoenix interaction term is significant, indicating that the Phoenix response function has a different slope than the other three cities.

Figure 2 shows the estimated response functions in each city for Model 2, as well as the underlying data.

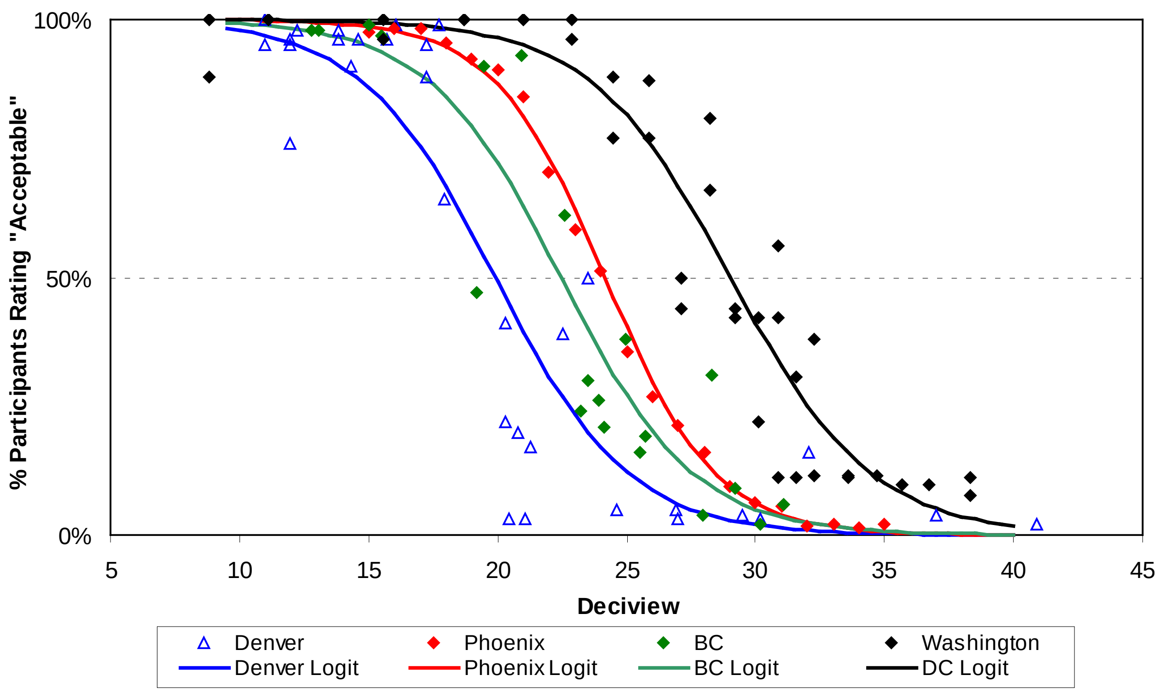

Figure 2. Estimated response functions for four cities using the fully interacted logit analysis.

The significantly different slope of the Phoenix response function is clearly visible in Figure 2. The negative estimated coefficient on the Phoenix interaction term results in the Phoenix response function being steeper than the other cities' functions. In other words, Phoenix respondents' acceptance probabilities were more sensitive to changes in dv levels. Figure 2 also shows the Washington, DC function is modestly less steep than the others, but the decrease in the slope is not statistically significant. Therefore, while Washington, DC respondents are more likely to accept worse visibility overall, they are just as responsive to changes in dv as respondents in Denver and British Columbia.

As with Model 1, it is possible to use the Model 2 results to estimate the dv values where the estimated response functions cross the 50% acceptability level, as well as any alternative criteria levels. The Model 2 estimates of alternative acceptability dv values for each city are shown in Table 4.

Table 4. Model 2 estimated haziness (dv) levels of alternative acceptability criteria

	Denver	British Columbia	Washington, DC	Phoenix
90% acceptability criteria	14.21	16.80	23.03	24.15
75% acceptability criteria	17.05	19.63	26.03	21.80
50% acceptability criteria	**19.90**	**22.45**	**29.12**	**24.15**
25% acceptability criteria	22.74	25.28	32.03	26.51
10% acceptability criteria	25.59	28.10	35.03	28.87

The Model 2 estimates of the 50% acceptability criteria are nearly identical to the Model 1 estimates; the biggest difference is a 0.07 dv decrease in the Washington, DC 50% acceptability criteria. The essentially identical estimates of the 50% acceptability criteria in Models 1 (city indicator only) and Model 2 (full interaction) indicates the choice of model form does not change the conclusion that the logit results are consistent with the 20 to 30 dv CPL range identified in the draft UFVA (EPA, 2009).

We also conducted hypothesis testing with the four city data used in this section to examine the probability that the 50% acceptance criteria in the four different cities are the same. We used the full interaction model results for the hypothesis testing. Our approach estimated the mean 50% criteria dv levels and standard error (based on the Krinsky-Robb confidence intervals) for each of the four cities. We then conducted a hypothesis testing using a t-test to estimate the probability the mean 50% criteria dv levels are the same in each pair of cities. The null hypothesis in this hypothesis test is that the means are the same. As shown in Table 5, the null hypothesis is strongly rejected for all pairs of cities, indicating that the mean 50% criteria dv levels differ for all four cities.

Table 5. Hypothesis testing on whether the full interaction model mean 50% criteria dv levels are the same

	British Columbia Mean dv = 22.45	Phoenix Mean = 24.15	Washington, DC Mean dv = 29.12
Denver Mean dv = 19.90	t-stat = 16.89 Pr(Den = BC) ~0	t-stat = 35.15 Pr(Den = Ph) ~ 0	t-stat = 30.21 Pr(Den = DC) ~ 0
British Columbia	–	t-stat =12.08 Pr(BC = Ph)~0	t-stat = 21.23 Pr(BC = DC) ~ 0
Phoenix		–	t-stat = 16.53 Pr(Ph = DC) ~ 0

Analysis of Washington, DC Preference Studies

There are two related studies of visibility preferences in Washington, DC. In 2001, in a project sponsored by the U.S. Environmental Protection Agency, Abt Associates conducted a pilot focus group study (DC-2001) of urban visibility preferences in Washington, DC. In 2009, in a study for the Utility Air Regulatory Group, Smith and Howell conducted a series of three tests of urban visibility preferences in Washington, DC. In their first test (DC-Test 1), Smith and Howell used all the images used in the DC-2001 study, trying to replicate the DC-2001 study. Their second test (DC-Test 2) used fewer of the Washington images, restricting the study to the 12 images with better visibility (images with visibility of 27.1 dv or better). In the third test (DC-Test 3), they expanded the range of images to include two hazier images (adding a 42 and 45 dv images, and deleting images at 11.1, 15.6, and 24.5 dv).

An important question is whether the participant responses obtained in the DC-2001 study are similar to the responses in Test 1, which was designed to replicate the DC-2001 study. A related question is whether the responses in Tests 2 and 3 are similar to Test 1. To investigate these questions we estimated logit response functions using the data from the four different Washington, DC data sets (DC-2001, DC-Test 1, DC-Test 2, and DC-Test 3), using the full interaction logit model specification.

The estimated coefficients from a full interacted model are presented in Table 6. The DC-2001 test is used as the omitted interaction variable.

Table 6. Logit regression results with full interacted model of Washington, DC studies

Variable	Coefficient (β)	Standard error	z-statistic	Pr |β| = 0	5% confidence estimate	95% confidence estimate
dv	-0.4035	0.0567	-7.12	< 0.001	-0.5146	-0.2925
Test 1	-1.5425	1.8785	-0.82	0.412	-5.2242	2.1392
Test 2	-0.7431	2.0737	-0.36	0.720	-4.8075	3.3212
Test 3	3.4109	2.6980	1.26	0.206	-1.8772	8.6990
Test 1 × dv	0.0616	0.0632	0.97	0.330	-0.0624	0.1855
Test 2 × dv	-0.1043	0.0804	-1.30	0.194	-0.2618	0.0532
Test 3 × dv	-0.0607	0.0868	-0.70	0.485	-0.2309	0.1095
Constant	11.5621	1.6777	6.89	< 0.001	8.2739	14.8504

Figure 3 shows the estimated full interaction logit function for the separate Washington, DC Test data, including the DC-2001 data.

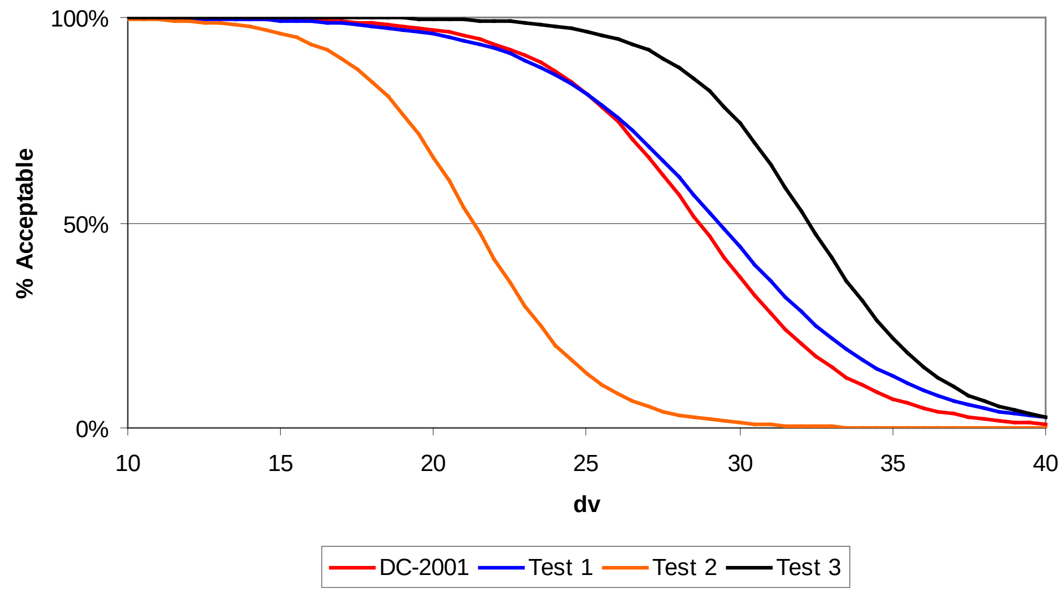

Figure 3. Full interaction logit results for four separate DC data sets.

Figure 3 suggests that while the logit functions from DC-2001 and Test 1 appear to be similar, Test 2 and Test 3 appear to be substantially different. Estimating the 50% criteria levels, along with the Krinsky-Robb confidence intervals, confirms this observation. Table 7 presents the estimated mean 50% criteria levels and the Krinsky-Robb confidence intervals.

Table 7. Mean 50% criteria levels, and Krinsky-Robb confidence intervals

	Mean dv	5% confidence estimate	95% confidence estimate
Test 1	29.30	28.59	29.97
Test 2	21.30	20.57	22.03
Test 3	32.26	31.37	33.16
DC-2001	28.65	27.46	29.70

Hypothesis testing using the predicted mean dv distribution from the Krinsky-Robb procedure provides statistical support for the conclusion that the Test 1 and DC-2001 results are similar, but Test 2 and Test 3 results are different. The hypothesis testing results are presented in Table 8.

Table 8. Hypothesis testing on the individual coefficients in the full interaction model is the same for the four different Washington, DC experiments

	Test 2 **Mean dv = 21.30**	**Test 3** **Mean = 32.26**	**DC-2001** **Mean dv = 28.65**
Test 1 Mean dv = 29.30	Reject hypothesis that Test 1 = Test 2 (pr < 0.001)	Reject hypothesis that Test 1 = Test 3 (pr < 0.001)	Cannot reject hypothesis that Test 1 = DC-2001 (pr = 0.15)
Test 2	–	Reject hypothesis that Test 2 = Test 3 (pr < 0.001)	Reject hypothesis that Test 3 = DC-2001 (pr < 0.001)
Test 3		–	Reject hypothesis that Test 3 = DC-2001 (pr < 0.001)

As shown in Table 8, we cannot reject (at the 5% confidence level) the hypothesis that the mean 50% criteria level in the DC-2001 data and the Test 1 data are the same. In other words, it is likely that the mean dv in Test 1 is the same as the mean dv in DC-2001. Thus, this hypothesis test supports combining those two data sets together, as we did in the four city analysis presented above. The results in Table 8 reject the hypothesis that Test 2 and Test 3 are the same as either the Test 1 or DC-2001 results.

Further Analysis of the Washington, DC Test 1 Data

Smith and Howell conducted Test 1 using three distinct groups of respondents. Four of the respondents in Test 1 were Washington, DC area residents that were used in a pilot test of the testing procedure. Twelve of the respondents were CRA International employees who live in the Washington, DC area, and ten of the respondents were CRA International employees who live in the Houston, Texas area. The Test 1 participants were all shown the same images of Washington, DC haze levels as the DC-2001 participants, and were asked about their preferences for urban visibility in Washington, DC.

We investigated heterogeneity among these three groups' responses by conducting a full interaction logit analysis using information about which of the three groups (pilot, DC area or Houston area) the respondents were in. We also included the DC-2001 respondents (who were all DC area residents) in this analysis to conduct hypothesis tests on whether the Test 1 groups were different than the DC-2001 respondents. We used the pilot test respondents as the omitted group in a full interaction model analysis. The results of the logit analysis are presented in Table 9.

Table 9. Logit regression results with full interacted model of the 3 Test 1 groups and the DC-2001 participants

Variable	Coefficient (β)	Standard error	z-statistic	Pr $\|\beta\|$ = 0	5% confidence estimate	95% confidence estimate
dv	-0.5719	0.1310	-4.36	0.000	-0.8287	-0.3151
Test 1/DC	-0.8344	3.7361	-0.22	0.823	-8.1570	6.4881
Test 1/ Houston	-4.8831	3.5486	-1.38	0.169	-11.8382	2.0719
DC-2001	-2.4042	3.7273	-0.65	0.519	-9.7095	4.9012
Test 1/DC × dv	0.1439	0.1420	1.01	0.311	-0.1344	0.4222
Test 1/Houston × dv	0.2643	0.1372	1.93	0.054	-0.0047	0.5332
DC-2001 × dv	0.1684	0.1428	1.18	0.238	-0.1114	0.4482
Constant	13.9663	3.3284	4.20	0.000	7.4428	20.4898

Using the estimated coefficients in Table 9, we calculated estimated 50% criteria levels for each group, along with the Krinsky-Robb confidence intervals, which are shown in Table 10.

Table 10. Mean 50% criteria levels, and Krinsky-Robb intervals for the Test 1 groups and the DC-2001 participants

	Mean dv	5% confidence level	95% confidence level
Test 1/DC	30.68	29.79	31.51
Test 1/Houston	29.52	28.30	30.66
Test 1/Pilot	24.42	22.37	25.97
DC-2001	28.65	27.46	29.70

Table 10 suggests that the mean 50% acceptance criteria level for the Washington, DC area residents in the 2001 study are closest to the mean 50% criteria level for the Test 1 Houston area residents, and differ to a greater degree from the mean 50% criteria level for the Test 1 Washington area residents. Hypothesis testing confirms this finding, as shown in Table 11.

Table 11. Hypothesis tests of the mean 50% acceptance criteria level for the three groups in the Test 1 data and the DC-2001

	Houston Mean dv = 29.52	**Pilot Mean dv = 24.42**	**DC-2001 Mean dv = 28.65**
Test 1/DC area Mean dv = 30.68	Reject hypothesis Houston = Test 1/DC (pr = 0.06)	Reject hypothesis Pilot = Test 1/DC (pr < 0.001)	Reject Test 1/DC = 2001-DC (pr < 2%)
Test 1/Houston area	–	Reject Houston = Pilot (pr < 0.001)	Cannot reject Houston = DC-2001 at 5% confidence (pr = 14%)
Test 1/Pilot		–	Reject Pilot = DC-2001 (pr <0.001)

These hypothesis test results in Table 11 provide some insight into the hypothesis tests in Table 8, which found the 50% mean criteria level (mean = 29.30 dv) estimated using the combined Test 1 data is similar to the 50% criteria level from the DC-2001 data (mean = 28.65 dv). The Table 11 results suggest that the Table 8 results could be the result of the Test 1 pilot participants (mean = 24.42 dv) offsetting the Test 1/DC area participants (mean = 30.68 dv), giving us a mean estimate for the combined sample closest to the Houston area participants (mean = 29.52 dv).

Individual Heterogeneity

Individual respondents will likely have different general attitudes regarding haze than other respondents, reflecting their individual preferences about urban visibility. An individual's preferences may affect how they rate the acceptability of different dv levels. In the Smith and Howell (2009) Washington, DC study we can track an individual's responses over all dv levels.[3] This enables us to account for individual heterogeneity in our estimation procedure using individual-specific indicators. These are called fixed-effect models and control for unobserved differences between respondents.

We conducted a logit analysis on Test 1 data using individuals as the indicator variable. We included slope interaction terms for the Washington and Houston area residents (with the pilot slope interaction term omitted). Each individual[4] also has an indicator which becomes the

[3] While this level of data was originally collected for the studies in Denver, Phoenix, British Columbia and the 2001 Washington, DC study, the original data is not available at this time.

[4] Respondents 1 and 13 are dropped in the individual heterogeneity analysis because they had identical responses, accepting every dv level. The form of the logit model used in this analysis cannot be estimated when all the responses are identical.

intercept term for that individual. The terms for Respondents 2 through 12 are intercept shifters for DC respondents. Respondents 14 through 22 were Houston respondents, and Respondents 23 through 25 were pilot respondents. The results from this model are presented in Table 12.

Table 12. Logit analysis results of individual heterogeneity analysis

Variable	Coefficient (β)	Standard error	z-statistic	Pr |β| = 0	5% confidence estimate	95% confidence estimate
dv	-0.7315	0.1911	-3.83	0	-1.1060	-0.3569
Houston × dv	0.1207	0.2139	0.56	0.573	-0.2986	0.5399
DC × dv	-0.3588	0.2658	-1.35	0.177	-0.8799	0.1622
Respondent 2 (DC)	35.2050	5.9847	5.88	0	23.4752	46.9349
Respondent 3 (DC)	35.2050	5.9847	5.88	0	23.4752	46.9349
Respondent 4 (DC)	29.9950	5.2578	5.7	0	19.6900	40.3001
Respondent 5 (DC)	34.3924	5.8635	5.87	0	22.9002	45.8846
Respondent 6 (DC)	32.0347	5.5755	5.75	0	21.1070	42.9624
Respondent 7 (DC)	31.0845	5.4326	5.72	0	20.4369	41.7322
Respondent 8 (DC)	25.7365	4.5956	5.6	0	16.7293	34.7438
Respondent 9 (DC)	36.1200	6.1617	5.86	0	24.0434	48.1966
Respondent 10 (DC)	34.3924	5.8635	5.87	0	22.9002	45.8846
Respondent 11 (DC)	28.7572	5.0615	5.68	0	18.8369	38.6775
Respondent 12 (DC)	35.2050	5.9847	5.88	0	23.4752	46.9349
Respondent 14 (H)	16.6104	2.8047	5.92	0	11.1133	22.1075
Respondent 15 (H)	15.9236	2.7170	5.86	0	10.5984	21.2488
Respondent 16 (H)	15.9236	2.7170	5.86	0	10.5984	21.2488
Respondent 17 (H)	18.3145	2.9999	6.11	0	12.4348	24.1942
Respondent 18 (H)	20.8740	3.3153	6.3	0	14.3761	27.3719
Respondent 19 (H)	13.3405	2.3443	5.69	0	8.7457	17.9353
Respondent 20 (H)	19.8140	3.1722	6.25	0	13.5966	26.0315
Respondent 21 (H)	16.6104	2.8047	5.92	0	11.1133	22.1075
Respondent 22 (H)	18.8166	3.0538	6.16	0	12.8313	24.8019
Respondent 23 (P)	16.0044	4.3526	3.68	0	7.4736	24.5353
Respondent 24 (P)	17.1746	4.6884	3.66	0	7.9854	26.3637
Respondent 25 (P)	19.9838	5.4132	3.69	0	9.3742	30.5933
Respondent 26 (P)	18.2301	4.9705	3.67	0	8.4882	27.9720

As in the analyses previously described, we used the logit analysis coefficients in Table 12 to estimate the mean value for the 50% acceptance criteria. We also estimated the Krinsky-Robb confidence intervals for each data subset using the fixed effects model. Because three of the Test

1 participants were deleted in the individual heterogeneity analyses, for comparison purposes we also re-estimated a model without accounting for individual heterogeneity using the same data set (i.e., with the two individuals deleted). The results are presented in Table 13.

Table 13. Estimated mean 50% criteria levels, and Krinsky-Robb intervals for the Test 1 data accounting for individual heterogeneity

	Mean dv	Lower bound 95%	Upper bound 95%
Washington area residents	30.57	29.97	31.18
Houston area residents	29.40	28.41	30.33
Pilot (DC residents)	24.40	22.60	25.91
Mean dv estimates without individual heterogeneity (using same data)			
Washington area residents	30.02	29.19	30.77
Houston area residents	28.50	27.25	29.58
Pilot (DC area residents)	24.42	22.37	25.97

Table 13 shows that including individual heterogeneity in the model modestly increased the estimated mean 50% criteria levels.

Table 14 shows the results of hypothesis testing on the individual heterogeneity results. Modeling with individual heterogeneity leads to rejecting the hypothesis that the mean dv levels are the same in any of the three respondent groups.

Table 14. Hypothesis tests of the mean 50% acceptance criteria level for the three groups in the Test 1 data modeled with individual heterogeneity

	Houston area Mean dv = 29.40	Pilot (DC area) Mean dv = 24.40
DC area Mean dv = 30.57	Reject hypothesis Houston = DC (pr = 0.02)	Reject hypothesis Pilot = DC (pr < 0.001)
Houston area	–	Reject Houston = Pilot (pr < 0.001)

Summary

This memorandum describes a series of logit regression analyses that estimated the percentage of respondents that rated a haze (dv) level acceptable in four different studies of urban visibility. The first analysis in this report estimated a separate logit function for each of the four studies: Denver, British Columbia, Phoenix and Washington, DC (combining the data from the DC-2001 study and all Test 1 data from the DC-2009 study). The estimated mean 50% criteria levels in the four cities (Table 4) are different, with the mean estimate ranging from 19.90 dv (Denver) to 29.03 dv (Washington, DC). The hypothesis tests presented in Table 5 found that there is a statistically different logit function in each city (rejecting the null hypothesis that there was a single function that applies to more than one city). The range of mean estimates from the 4 city logit analysis is similar to the Candidate Protection Level range of 20 dv to 30 dv described in the draft UFVA (EPA, 2009).

The remainder of this memorandum examined in more detail the data from the two Washington, DC studies. In the first analysis focusing on only the Washington, DC data, we compared the estimated mean 50% criteria levels from the 2001 study to the mean estimates from each of the three tests in the 2009 study. Figure 3 and Table 7 show the estimated mean levels in the 2001 (mean = 28.65 dv) and 2009, Test 1 (mean = 29.30 dv) studies were similar, while the Test 2 (21.30 dv) and Test 3 (32.26) mean levels were quite different. The hypothesis tests presented in Table 8 support that overall observation. The only hypothesis not rejected was the hypothesis that the DC-2001 and Test 1 are the same (i.e., we cannot reject the hypothesis that they have the same mean 50% criteria level). This finding supports our approach of combining the DC-2001 and the DC-2009, Test 1 data in the four city analysis.

In the second analysis of the Washington, DC data, we investigated whether the study participants who lived in the Washington, DC Metro area had the same mean 50% criteria levels as the participants who lived in the Houston metro area. This analysis involved three groups of Washington, DC residents (the DC-2001 participants, the pilot project participants in the DC-2009 study, and participants 1 through 12 in Test 1 of DC-2009). The hypothesis tests results in Table 11 show that the participants in the DC-2001 and the Houston area residents in the DC-2009 study are similar (i.e., we cannot reject the hypothesis they have the same mean 50% criteria level). Our hypothesis testing further found however, that the DC-2001 participants had statistically significantly different mean 50% criteria levels than either of the two groups of Washington, DC area residents included in the DC-2009, Test 1 results.

The third analysis of the Washington, DC data investigated the effect of individual heterogeneity of preferences. This analysis was limited to the DC-2009 data because it required more complete information on the responses of each participant. The individual heterogeneity analysis found modestly higher mean 50% criteria levels than the second analysis of the Washington, DC area

residents. The hypothesis testing in this analysis rejected the hypothesis that the mean dv levels were the same for the three groups who participated in Test 1.

This apparent inconsistency with the two hypothesis tests of analyses of subsets the Washington, DC studies with the results of the hypothesis tests comparing the DC-2001 data with all of the DC-2009, Test 1data may be due to having subdivided the participants of Test 1 into subsets with too few members to provide stable results. Combining the DC-2001 data with all the Test 1 data provide the largest sample size available to estimate the logit preference function for Washington, DC.

References

Hole, A.R. 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics* 16:827–840.

Krinsky, I. and L. Robb. 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics* 68(4):715–719.

Smith, A.E. and S. Howell. 2009. An Assessment of the Robustness of Visual Air Quality Preference Study Results. CRA International, Washington, DC. March 30. Prepared for the Utility Air Regulatory Group. Submitted as supplemental material to presentation by Anne Smith to the public meeting of the EPA Clean Air Science Advisory Council. April 2.

Stratus Consulting. 2009. Review of Urban Visibility Preference Studies. Prepared for Vicki Sandiford, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Stratus Consulting Inc., Boulder, CO. September 21.

U.S. EPA. 2009. Particulate Matter Urban-Focused Visibility Assessment; External Review Draft. U.S. Environmental Protection Agency. September.