

**RESPONSE TO COMMENTS**

COMMUTER CONNECTIONS – 2013 STATE OF THE COMMUTE SURVEY  
Draft Technical Survey Report  
September 17, 2013

Note – The November 19, 2013 draft of the SOC Survey Report incorporates small wording changes and a number of adjustments to fix typos and to clarify language in response to comments. The discussion below is to respond to other comments. Additionally, a section was added to capture respondents' use of primary roads used to get to work.

**Question – Representativeness of the sample with respect to income** - *"The survey is described as a statistically accurate depiction of a representative sample of commuters in the region. However, the report indicates that almost half (49%) of respondents have incomes of \$120,000 or more (p. 9). This is a surprisingly high percentage. By comparison, the 2011 American Community Survey reported that 26.5% of full-time, year-round workers with earnings in the DC-VA-MD Urbanized area earned \$100,000 or more (see attached). If only 26.5% of workers earn over \$100,000, it is difficult to understand how 49% of commuters can earn more than \$120,000."*

**Response** – The SOC survey collects data on household income, while the income distribution reported by the commenter is for earnings of individual workers. Since the Washington region includes many households with two or more workers, the household income, thus the two income distributions will not be the same.

**Question - Statistical validity of some conclusions** - The commenter says, "There are several instances where statements in the report do not appear to be supported by its findings, when the survey results are carefully examined from a statistical standpoint. In general, as the sample size decreases, the range in which the "true" or "accurate" answer increases." The commenter then references some basic principles of sample sizes and confidence levels that were provided in the 2007 State of the Commute Survey USERS GUIDE that was prepared for jurisdiction staff who received SOC data files for their individual use in analysis.

The commenter then says, "A spot-check of several statements in the report appears to indicate that several of them cannot be supported by the data provided in the tables that accompany them. He then provides specific examples in which the differences in percentages appear to be too small to be statistically different for the given sample sizes and says, "From just these two examples, it appears that all of the statements in the report should be verified for statistical validity at the implied 95% confidence level; one possible approach might be to lower the confidence level for the results for some sub-groups."

**Response** - The information provided in the USERS GUIDE was to serve simply as a rule of thumb to alert the data users, many of whom might be unfamiliar with statistical testing concepts, of this basic principle. But statistical difference should be tested using a chi-square test, since the pure "plus or minus" percentage result is strictly accurate only when results are

approximately 50%-50%. In the SOC survey, all results that are reported as statistically significant were analyzed using this test.

**Question - Page 8, first paragraph** – *The commenter asks how the expansion of the data ensure representativeness of the sample for demographic characteristics in "each of the 11 study areas."*

**Response** - The SOC survey is not a census, so the distributions for various demographic questions are subject to statistical variability. As mentioned on page 5 and described more fully in Appendix 1, the expansion process applied a correction for imbalances in ethnicity at the individual jurisdiction level. No additional corrections were made for age or income, for two reasons. First, income and ethnicity are interrelated, so correcting for ethnicity also corrects somewhat for income bias in the sample. Second, if weighting factors were to be developed to adjust simultaneously for ethnicity, income, and age, the resulting weighting factors would be derived from extremely small samples of respondents, leading to possible errors introduced through the weighting process.

**Question – Page 9, Figure 3** – *'Over 6300 interviews were conducted but apparently only 4439 (70%) responded to the household income question. How confident can we be that the reported income levels are representative of all respondents?'*

**Response** – The short answer is that we cannot know if the income levels are either completely accurate or completely representative. Respondents might omit the income question because they don't know their household's income, don't know it accurately enough to respond, or because they want to keep that information private. But income non-response is typically in the range of 20% to 40% of total respondents – even the U.S. Census non-response rate is about 20%. And studies conducted by the U.S. Census have shown that income non-response occurs across all income types, with slightly higher non-response rates at higher income levels. Some surveys, for which income is an important analysis variable, impute income for non-respondents, however this is a complex modeling exercise that is beyond the scope of the SOC survey. Additionally, while some SOC results are examined by income level, this is not a major element of the survey, so the reported distribution of income is assumed to be both as accurate and representative as can be collected and as is needed for this purpose.

**Question – Page 27, Table 12** – *The commenter said, "Suggest showing speeds that result from this table: commuter rail = 31 mph; CP = 27.6; DA=33; Bus=16.4; Metrorail=17.0; Bike=12.5; Walk=3.8"*

**Response** – The suggested addition was not added because it would show an "apples to oranges" comparison in which transit and alternative mode "speeds" would include both travel and access/wait time, while driving, bike, and walk would show only travel time. The speed does not seem to provide relevant and useful information.

**Question – Page 32, 2nd paragraph** – *The commenter said, "Questionable statistical significance between adjacent pairs dues to small sample size. (686 sample yield approx. ± 3.9% range.)"*

**Response** – The confidence level is not relevant in this instance. The paragraph is simply reporting the distribution of the responses for the 686 respondents. The percentages are not being compared across sub-groups.

**Question – Page 54, First paragraph** – What is the source of the 54% in the paragraph about HOV lane use and influence?

**Response** – This figure is not shown in any table, nor is it in Figure 36. It is the response to a “yes or no” question asking if the use of the HOV lane influenced their decision to use an alternative mode. It is the share of respondents who answered yes. Figure 36 compares the mode shares for all respondents who said HOV lanes were available and those who said HOV lanes were not available.

**Question – Page 101, Sixth bullet** – The commenter asks where in the main document that the percentages are shown?

**Response** – The bus distance is noted on page 47, just above Figure 30. The train distance is in the paragraph below Figure 30.