Appendix F: Resident Survey Data Analysis Methodology

The first part of the analysis of the resident survey was based on pair-wise comparisons of survey results, comparing each center with the control group values. After that, regression analysis was used to control for demographic characteristics that vary across individual survey respondents.

Pair-wise Comparisons

Much of the analysis used comparisons across the centers and the control group. All such comparisons were pair-wise, comparing each center (divided into inner and outer rings) with the control group. The graphs in the report indicate when values for a center are statistically significantly different from levels from the control group. The test of statistical significance is used to measure whether the differences in the survey data (centers versus the control group) arise from differences in the underlying populations, and so would be regarded as real differences, or are simply due to sampling error. The graphs indicate when values differ at the five percent level, meaning that the probability that the difference is due to sampling error is five percent or less. This is the most common measure of statistical significance.

As an example, consider the data for commute mode for the outer ring of Torrance and the Pacific Coast Highway control group. In the outer ring of Torrance, 88.12% of respondents commute to work by car, compared to 94.81% of respondents who commute to work by car in the control group. That difference is statistically significant at the five percent level, indicating that the difference is likely to be due to real differences in commuting in the full population in those areas rather than sampling error. Stated more formally, there is less than a five percent chance that the difference in car commuting percentages across outer ring Torrance and the control group is due to sampling error. The formulae for tests of statistical significance used in the analysis are given below.

Tests of differences between sample proportions:

$$\frac{p_1 - p_2}{\sqrt{\frac{p_1(100 - p_1)}{n_1} + \frac{p_2(100 - p_2)}{n_2}}}$$

where p_1 = sample proportion from sample number 1 (i.e. percent who commute by car in outer ring Torrance in above example)

p_2 = sample proportion from sample number 2 (i.e. percent who commute by car in control group in above example)

n_1= sample size from sample number 1
n_2 = sample size from sample number 2

For variables that would be expressed as sample averages, such as numbers of trips, the below formula to test for the significance of differences between sample means was used.

$$\frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where m_1 = sample mean for sample 1
m_2 = sample mean for sample 2
σ_1 = sample standard deviation for sample 1
σ_2 = sample standard deviation for sample 2
n_1 = sample size for sample 1
n_2 = sample size for sample 2

Regression Analysis

Comparing travel behavior across centers and the control group provides insights, but one difficulty is that the demographic characteristics of individuals varies across centers and between pairs composed of a center and the control group. Thus pair-wise comparisons do not give definitive information about the role of urban design as opposed to differing characteristics of individuals who live in the centers and control group area. To address that issue, we ran regression analyses to explain the number of trips an individual took during the one-day diary period as a function of that individual's demographic characteristics. The regressions also included dummy variables that took on values of zero or one to indicate the person's neighborhood of residence. Two regressions were analyzed – one for total driving trips and one for total walking trips. For both driving and walking trips, the totals are the total individual trips made by a survey respondent in one day.

The results for the ordinary least squares (OLS) regression for total driving trips are shown below.

Regression Results, Number of Total Driving Trips

	coefficient	standard error	t-statistic	prob > |t|
age1825	0.645	0.542	1.190	0.234
age2640	**0.852**	**0.302**	**2.821**	**0.005**
age4165	**1.148**	**0.267**	**4.300**	**0.000**
Female	-0.030	0.195	-0.154	0.879
incunder15	**-1.593**	**0.573**	**-2.780**	**0.006**
inc1535	-0.842	0.439	-1.918	0.055
inc3555	-0.311	0.283	-1.099	0.271
inc5575	0.208	0.252	0.825	0.410
Ingi	-0.932	1.015	-0.918	0.359
Ingo	**-1.401**	**0.370**	**-3.786**	**0.000**
Rivi	-0.330	0.324	-1.019	0.309
Rivo	**-0.632**	**0.260**	**-2.431**	**0.015**
Toti	-0.284	0.384	-0.740	0.459
Toto	-0.263	0.283	-0.929	0.353
Constant	2.326	0.295	7.885	0.000
N	657			
R-squared	0.0821			
R-squared adj	0.0621			

Variables are dummy variables (zero or one) to indicate gender, age ranges, and income:

age1825 = 1 if respondent is between 18 and 25 years
age 2640 = 1 if respondent is between 26 and 40 years
age4165 = 1 if respondent is between 41 and 65 years
age over 65 is the omitted age category
incunder15 = 1 if respondent reports income less than $15,000 per year
inc1535 = 1 if respondent reports income between $15,001 and $35,000 per year
inc3555 = 1 if respondent reports income between $35,001 and $55,000 per year
inc5575 = 1 if respondent reports income between $55,001 and $75,000 per year
income greater than $75,000 per year is the omitted income category
female = 1 for women
Dummy variables = 1 for residence in Inglewood inner ring (INGI), Inglewood outer ring (INGO), Riviera Village inner ring (RIVI), Riviera Village outer ring (RIVO), Torrance inner ring (TOTI), and Torrance outer ring (TOTO). The control group, Pacific Coast Highway, is the omitted category.

To allow comparisons across center locations, the coefficients from the regression were used to calculate predicted daily trip generation rates for a woman aged 26-40 years with annual income between $55,001 and $75,000. The dummy variables show the magnitude of the difference in driving trips between centers and the Pacific Coast Highway control group, controlling for the gender, age, and income variables. The difference (from the control group) in driving trip generation is statistically significant for Inglewood outer ring and Riviera Village outer ring. The differences are meaningfully large in magnitude, as discussed in the text of the report.

Some individuals did not take any trips during the diary period, making the dependent variable a combination of a discrete and linear variable. For that reason, a tobit regression was also used, with the same independent variables as appear in the table above. The pattern of sign and statistical significance for the neighborhood dummy variables did not change using the tobit regression, and the magnitude of driving trip reduction was similar using the tobit regression. For ease of interpretation, the OLS results are shown above.

The same analysis was repeated for walking trip generation – i.e. the dependent variable was the number of walking trips made by an individual during the one-day travel diary period. The results are shown below.

Regression Results, Number of Total Driving Trips

	coefficient	standard error	t-statistic	Prob > |t|
age1825	-0.001	0.144	-0.007	0.994
age2640	**0.214**	**0.080**	**2.675**	**0.008**
age4165	**0.207**	**0.071**	**2.915**	**0.004**
female	0.007	0.052	0.135	0.898
incunder15	-0.016	0.153	-0.105	0.915
inc1535	0.203	0.117	1.735	0.083
inc3555	-0.010	0.073	-0.137	0.893
inc5575	0.060	0.067	0.896	0.371
ingi	-0.052	0.270	-0.193	0.848
ingo	-0.147	0.098	-1.500	0.136
rivi	**0.243**	**0.086**	**2.826**	**0.005**
rivo	0.082	-0.069	-1.188	0.239
toti	0.097	0.102	0.951	0.341
toto	-0.0001	0.075	-0.001	0.998
constant	-0.033	0.078	-0.423	0.671
N	657			
R-squared	0.042			
R-squared adj	0.021			

As was the case for driving trip generation, using tobit instead of OLS for the regression for walking trips produced no changes in sign or significance of coefficients. For both regressions above, the R-squared statistics are in the range of what is common when explaining individual trip generation.[1]

[1] See, e.g., Kitamura, Ryuichi, Patricia Mokhtarian, and Laura Laidet, "A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area," *Transportation* 24: 125-158, 1997, who in a

survey study that included travel diary data for households, found that regressions explaining the number of non-motorized trips made by diary respondents had R-squared values ranging from 0.0256 to 0.0428. Similarly see Greenwald, Michael and Marlon G. Boarnet, "The Built Environment as a Determinant of Walking Behavior: Analyzing Non-Work Pedestrian Travel in Portland, Oregon," *Transportation Research Record*, number 1780, pp. 33-42, 2002, who in predicting individual non-work walking trips, found R-squared values from 0.0509 to 0.0848 when using ordinary least squares.