INTRODUCTION

The study team completed a number of tasks as part of data processing. The primary goals of data processing tasks were twofold as noted below:

1. Identify and rectify potential accuracy and validity issues, and
2. Enhance the data for subsequent analysis.

This appendix describes the checks in support of these data processing goals. The checks performed can be categorized into three groups: Uniformity Checks, Consistency Checks, and Logical Checks. In addition, a list of checks that were typically carried out by the modeling consultant were also performed on the final dataset. Below, each of the four types of checks are described in greater detail. In the last section, the changes implemented in response to the checks are summarized.

UNIFORMITY CHECKS

The purpose of the uniformity checks was to ensure that the variables, and values contained in the data files followed assumed notations for variable names, variable descriptions, variable values, and value descriptions, among others. For example, ensuring the values match the responses enumerated in the questionnaire document. Furthermore, the checks also ensured that all data files and the information contained in them were uniformly organized. For example, ensuring that a value of -99 was used throughout to represent a "not applicable" and -999 is used to represent a "missing" value. Table H.1 provides a list of specific uniformity checks that were applied to all data files.

Table H.1: List of specific uniformity checks

Checked variable name for intuitiveness
Checked variable description and matched with questionnaire text and branching logic
Checked whether the variable type in SPSS matches the type of values elicited in the survey (i.e. string versus numeric)
Checked response values against questionnaire
Checked description of response values against questionnaire
Ensured that value -99 is used for "not applicable"
Ensured that value -999 is used for "missing" entries
Ensured plausibility of missing entries
Matched column type (and other attributes) for common variables across files

CONSISTENCY CHECKS

The purpose of the consistency checks was to ensure that the values reported for a given variable are consistent with information contained in another related variable in the same or different data files. For example, school type (schooltype) is only asked of adults (age >= 18) who are currently students (student= 1 or 2). As can be seen, school type and student status provide different items of information but they are related and thus can be compared to evaluate the information contained in the variables. Table H.2 provides the specific checks that were performed at the household level. Tables H.3 and H.4 provide the specific checks that were performed at the person level and trip level respectively. No consistency checks were performed for the vehicle information as there was no information in the dataset to compare against.

LOGICAL CHECKS

The purpose of the logical checks was to compare responses by using information contained in variables in the same data file or across other data files. For example, starting location type on assigned travel date (diary_st_loc) and ending location type on assigned travel date (diary_end_loc) are reported in the person file. This information can be matched against the origin and destination purpose for the first and last trip reported in the trip file. The focus of the logical checks was mainly on the travel responses[1]. The logical checks that were performed are enumerated in Table H.5.

ADDITIONAL CHECKS FROM MODELING CONSULTANT

These represent checks that are typically carried out by the modeling consultant as part of their model development efforts. A list of these checks at the household, person, and trip levels are presented in Table H.6, Table H.7 and Table H.8 respectively. It must be noted that there is an overlap between some of these checks and the three types of checks described above.

ACTIONS TAKEN TO FIX ISSUES AND ENHANCE DATA

While most of the checks did not result in any issues, some discrepancies were found. The study team rectified majority of these issues in the following ways. First, a new column was derived or calculated, based on the original or revised data column(s). Second, recoded column(s) were created by rectifying the original data column(s). As noted in Appendix K, all data columns are appropriately tagged to help identify the type of column. Also, when a data column was recoded, the original data column was retained to facilitate comparisons. Third, where appropriate, changes that have cascading impacts were incorporated. For example, change to destination purpose (d_purpose), will have implications for o_purpose and also for t_purpose. It must be noted that there may still be some discrepancies that either could not be rectified (e.g. day level HOV usage is at odds with trip level usage) or were not identified because the study team did not enumerate the check that would result in the said discrepancy.

---

[1] As it was observed that the household and person level consistency checks also cover logical checks for information contained in those files.

Table H.2: List of household specific consistency checks

Name of the variable being checked	Related variable or table against which check was made	Description of the check
Household file: hh_tripcount	Trip file	Compared against count of trip records in the trip file for each household Compared against aggregate of person level trip count variable: person_tripcount
Household file: transithh	Trip file: mode_category	Compared to see if members of the household reported at least one trip with mode_category = 6
Household file: recruit_duration	Household file: recruit_start_et, recruit_end_et	Compared against the difference
Household file: numvehicle_3plus	Household file: numvehicle	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: hhsize_5plus	Household file: hhsize	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: numvehicle	Person file	Compared against count of vehicle records in the vehicle file for each household
Household file: numpersons	Person file	Compared against count of person records in the person file for each household
Household file: numadults	Person file: age_broad	Compared against counts of persons in the person file where age_broad >= 5
Household file: numkids	Person file: age_broad	Compared against counts of persons in the person file where age_broad < 5
Household file: numkids_3plus	Household file: numkids	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: numstudents	Person file: schooltype	Compared against counts of persons in the person file where schooltype is any valid value except "None"
Household file: numstudents_adult	Person file: age_broad and schooltype	Compared against counts of persons in the person file where age_broad >= 5 and schooltype is any valid value except "None"
Household file: numstudents_kid	Person file: age_broad and schooltype	Compared against counts of persons in the person file where age_broad < 5 and schooltype is any valid value except "None"
Household file: numworkers	Person file: employment	Compared against counts of persons where employment = 1 or 2
Household file: numworkers_3plus	Household file: numworkers	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: numdrivers	Person file: license	Compared against counts of persons where license = 1
Household file:	Household file: hhincome_detailed	Compared to ensure that the mapping between the disaggregate and

Name of the variable being checked	Related variable or table against which check was made	Description of the check
hhincome_broad		consolidated variables was appropriate for all "prefer not to answer" at the disaggregate level.
Household file: hhincome_detailed	Household file: hhincome_followup	Compared to ensure that the counts of valid values of detailed income variable align with "not applicable" Also, compared to ensure that the counts of "prefer not to answer" of the detailed income variable match with a valid response for the income follow-up variable
Household file: imputed_income_for_missing	Household file: hhincome_detailed	Compared to ensure that the counts of "prefer not to answer" for detailed variable match counts of valid imputed values
Household file: imputed_hhincome_detailed	Household file: hhincome_detailed, imputed_income_for_missing	Compared to ensure that the original values and imputed values in the new variable match with the detailed income and imputed income columns
Household file: r_numvehicles_4plus	Household file: numvehicles	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: r_household_income_rev_2	Household file: imputed_hhincome_detailed	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: r_numworkers_3plus	Household file: numworkers	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: r_numpersons_6plus	Household file: numpersons	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Household file: r_household_type	Person file: relationship	Values from the relationship were used to create the three categories of the household type; in defining "family couple households" the reference person was used to establish the presence of a couple
Household file: r_presence_of_children	Household file: numkids	Compared to ensure that the variable assumes a value 1 whenever number of kids is greater than zero and a value zero otherwise; includes other children in the household who are not related also

Table H.3: List of person specific consistency checks

Name of the variable being checked	Related variable or table against which check was made	Description of the check
Person file: age	Person file: age_broad	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Person file: age	Person file: age_followup	Compared to ensure that the "Not Applicable" of age_followup and "Prefer not to answer" for age variables align
Person file: employment	Person file: age_broad	Confirmed the branching logic by checking whether valid values were available for all people with age_broad >= 4
Person file: jobs_count	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1 or 2
Person file: work_hours	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1 or 2
Person file: student	Person file: age_broad	Confirmed the branching logic by checking whether valid values were available for all people with age_broad >= 5
Person file: schooltype	Person file: age_broad, student	Confirmed the branching logic by checking whether valid values were available for all people with age_broad <5 and student = 2 or 3 Also, confirmed response values against the possible values for various groups of ages as noted in the questionnaire
Person file: education	Person file: age_broad	Confirmed the branching logic by checking whether valid values were available for all people with age_broad >= 5
Person file: license	Person file: age_broad	Confirmed the branching logic by checking whether valid values were available for all people with age_broad >= 4
Person file: transit_fare	Person file: transit_freq	Confirmed the branching logic by checking whether valid values were available for all people with tansit_freq not equal to "Never"
Person file: transit_pass	Person file: transit_fare	Confirmed the branching logic by checking whether valid values were available for all people with transit_fare as "Used pass (any type)"
Person file: transit_discount	Person file: transit_fare	Confirmed the branching logic by checking whether valid values were available for all people with transit_fare not equal to "Free (no cost)"
Person file: school_freq	Person file: student, age_broad, schooltype	Confirmed the branching logic by checking whether valid values were available for all children and adult school enrollees who are not home schooled (i.e. schooltype not equal to 2, 6 or 95) Also, confirmed response values against the possible values for various groups

Name of the variable being checked	Related variable or table against which check was made	Description of the check
		of ages as noted in the questionnaire
Person file: school_lat, school_lng	Person file: schooltype, school_freq	Confirmed to make sure that for all people with schoolype~= (2, 6, 95) and do not take online classes (i.e. school_freq ~= 8)
Person file: school_mode	Person file: student, age_broad, schoolfreq, license	Confirmed the branching logic by checking whether valid values were available for all children and adult school enrollees who do not take classes online Also, confirmed response values against the possible values for various groups of ages as noted in the questionnaire
Person file: work_lat, work_lng	Person file: employment, workplace	Confirmed to make sure that for all people with employment=1, 2 or 3 and a fixed workplace, a valid latitude and longitude values are available
Person file: workplace	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3
Person file: occupation	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3
Person file: industry	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3
Person file: commute_freq	Person file: workplace	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3 and workplace = 1, 2
Person file: commute_mode	Person file: workplace, license	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3 and workplace = 1 or 2 Also, confirmed response values against the possible values based on licensing status as noted in the questionnaire
Person file: work_flex	Person file: employment	Confirmed the branching logic by checking whether valid values were available for all people with employment = 1, 2 or 3
Person file: diary_duration	Person file: diary_start_et, diary_end_et	Checked against the difference between the start and end times
Person file: person_tripcount	Trip file	Compared against count of trip records in the trip file for each household
Person file: person_tripcount_10plus	Person file: person_tripcount	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Person file: trips_first, trips_last	Person file: copytrips_confirm	Compared against the copytrips_confirm indicator to ensure that they align
Person file: trips_yesno	Person file: person_tripcount	Compared against count of trips at the person level
Person file: no_travel_dayoff,	Person file: person_tripcount	Compared to see if valid values were available when person_tripcount = 0

Name of the variable being checked	Related variable or table against which check was made	Description of the check
no_travel_telecommute, no_travel_workhomenoapay, no_travel_kidsvacation, no_travel_notransport, no_travel_sick, no_travel_visitor, no_travel_other		
Person file: toll_use_travelday, hov_use_travelday, park_pay_use_travelday	Person file: age_broad, person_tripcount	Confirmed the branching logic by comparing against count of trips at the person level and the age_broad >= 4
Person file: r_age	Person file: age_broad	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Person file: r_gender	Person file: gender	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Person file: r_employment	Person file: employment	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate
Person file: r_mode_to_work	Person file: employment, commute_mode	Compared to ensure that the mapping between the disaggregate and consolidated variables was appropriate

Table H.4: List of trip specific consistency checks

Name of the variable being checked	Related variable or table against which check was made	Description of the check
Trip file: departure_time	Trip file: departure_time_hhmm	The two variables provide the same information in different formats. These were compared to ensure consistency in the variables
Trip file: arrival_time	Trip file: arrival_time_hhmm	The two variables provide the same information in different formats. These were compared to ensure consistency in the variables
Trip file: departure_hour	Trip file: departure_time_hhmm	Compared the value of departure hour derived from the complete departure time
Trip file: arrival_hour	Trip file: arrival_time_hhmm	Compared the value of arrival hour derived from the complete arrival time
Trip file: reported_duration	Trip file: departure_time, arrival_time	Checked against the difference between arrival and departure times
Trip file: travelers_total	Trip file: travelers_hh, travelers_nonhh	Compared against sum of household members and non-household members in the travel party
Trip file: travelers_hh	Trip file: hhmember1, hhmember2, hhmember3, hhmember4, hhmember5, hhmember6, hhmember7, hhmember8, hhmember9	Compared against sum of household member occupancy indicators.
Trip file: o_purpose	Trip file: d_purpose	Checked to see that o_purpose for all but the first trip for each person corresponds to d_purpose for previous trip
Trip file: mode_category	Trip file: mode	Checked to make sure that the definitions of the disaggregate and aggregate variables align
Trip file: mode_hhauto	Trip file: mode_category, vehnum	Checked to see that this aligns with mode_category = 1, 2, or 3 Also, checked to see if a valid household vehicle ID was reported
Trip file: tollroad_cost	Trip file: tollroad_use, tollroad_cost_dk	Checked to see that valid values were available for tollroad_use = 1 and tollroad_cost_dk = 0
Trip file: park_cost	Trip file: park_pay, park_cost_dk	Checked to see that valid values were available for park_pay = 1 and park_cost_dk = 0
Trip file: taxi_cost	Trip file: mode_category, taxi_cost_dk	Checked to see that valid values were available for mode_category = 5 or 6 and taxi_cost_dk = 0
Trip file: transit_access	Trip file: mode_category, mode	Checked to see that valid values are available for mode_category = 6 and mode not equal to 12 (i.e. only transit without paratransit)
Trip file: transit_egress	Trip file: mode_category, mode	Checked to see that valid values are available for mode_category = 6 and

Name of the variable being checked	Related variable or table against which check was made	Description of the check
		mode not equal to 12 (i.e. only transit without paratransit)
Trip file: transit_pay	Trip file: mode_category, mode	Checked to see that valid values are available for mode_category = 6 and mode not equal to 12 (i.e. only transit without paratransit)
Trip file: transit_cost	Trip file: transit_pay	Checked to see that valid values are available for transit_pay = 2 or 3
Trip file: transit_system_1, transit_system_2, transit_system_3, transit_system_4	Trip file: mode_category, mode	Checked to see that valid values are available for mode_category = 6 and mode not equal to 12 (i.e. only transit without paratransit)
Trip file: transit_line_1, transit_line_2, transit_line_3, transit_line_4	Trip file: mode_category, mode	Checked to see that valid values are available for mode_category = 6 and mode not equal to 12 (i.e. only transit without paratransit)
Trip file: transit_transfers	Trip file: mode_category, mode, transit_system_1, transit_system_2, transit_system_3, and transit_system_4	Checked to see that valid values are available for mode_category = 6 and mode not equal to 12 (i.e. only transit without paratransit) Also, checked to see that the number of transfers matches with the valid information contained in the transit_system_1, transit_system_2, transit_system_3, and transit_system_4
Trip file: hhmember_none	Household file: hhsize Trip file: travelers_hh	Checked to see if valid values exist only when the household size is greater than one Also, checked to see if response to the variable is "no" only when the travelers_hh is greater than 1
Trip file: gdistance_bins	Trip file: gdistance	Checked to make sure that the bin ranges and values match
Trip file: gduration_bins	Trip file: gduration	Checked to make sure that the bin ranges and values match

Table H.5: List of trip specific logical checks

Name of the variable being checked	Related variable or table against which check was made	Description of the check
Trip file: hov_use	Person file: hov_use_travelday Trip file: mode, travelers_total	Compared the trip level hov usage against the day level hov usage reported
Trip file: tollroad_use	Person file: toll_use_travelday Trip file: mode	Compared the trip level toll usage against the day level toll usage reported
Trip file: park_loc	Person file: park_pay_use_travelday Trip file: mode	Compared the trip level parking location against the day level parking pay reported
Trip file: park_pay	Person file: park_pay_use_travelday Trip file: mode, park_loc	Compared the trip level parking pay usage against the day level parking pay reported
Trip file: driver	Person file: license Trip file: mode	Expected valid values for mode = 3, 4, and 7 & travelers_total > 2
Person file: dirary_loc_start	Trip file: o_purpose	Compared against the origin purpose of the very first trip
Person file: diary_loc_end	Trip file: d_purpose	Compared against the destination purpose of the very last trip
Trip file: o_purpose	Trip file: d_purpose	Checked to see if the combinations of o_purpose and d_purpose make sense
Trip file: d_lat, d_lng	Household file: home_lat, home_lng Trip file: d_name	This was the first of checks attempting to confirm the location accuracy. Checked to see if the home locations are consistent with destination reported in the trip file for "Go Home"

Table H.6: List of household specific checks provided by the modeling consultant

Missing geocodes, home
Sign inconsistency in geocodes, home
Geocodes out of the region, home
Length of the household ID must be constant.
Missing household income
Number of people in the person file must equal the household size variable in the Household file
Households in the household file do not show up in the trip file (except for the households with zero trip)
Number of spouses is greater than one

Table H.7: List of person specific checks provided by the modeling consultant

Missing geocode, work
Sign inconsistency in geocode, work
Geocode out of the region, work
Missing geocode, school
Sign inconsistency in geocode, school
Geocode out of the region, school
Household id in person file not in household file
Person represented in the person-file, but not represented in the trip file
(except for person with zero trip)
Child under 16 has driver's license
Missing educational attainment
Unusual age for school type
School type missing from student
School type present in non-student
School address missing from student
Work status missing
Child under 16 is listed as working
Work responses missing from worker
Worker with fixed location missing address
Worker did not report occupation
Worker did not report work location status (home, other, varies)
Worker did not report flexible schedule
Worker did not report work hours
Worker did not report industry
Missing school status
Missing gender
Missing age
Missing license
Missing relation to household head

Table H.8: List of trip specific checks provided by the modeling consultant

Missing geocode, origin
Missing geocode, destination
Sign inconsistency in geocode, origin
Sign inconsistency in geocode, destination
Duplicate trip record
Name/origin activity/type do not match information from the previous destination
Origin latitude/longitude do not match information from the previous destination
Missing time information
Arrival time is after departure time
Departure time is greater than the preceding trip's arrival time
Auto trip is missing passenger/driver info
Missing party size in auto trip
Missing/invalid household members in party in auto trip
Household members in party exceeds household size
Household members in party exceeds party size
Missing origin or destination name
The start time on the diary is not 3 AM
Origin matches destination by latitude/longitude