THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

PRODUCT 0-7054-P3
TxDOT PROJECT NUMBER 0-7054

Incorporation of Stated Preference and Revealed Preference Methods in Regional Travel Survey Programs: Final Guidebook

Katherine Asmussen
Aupal Mondal
Lisa Macias
Chandra Bhat

July 2022

Published October 2022

https://library.ctr.utexas.edu/ctr-publications/0-7028-P3.pdf

THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

Integration of Stated Preference and Revealed Preference Methods in Regional Travel Survey Programs: Final Guidebook

Katherine Asmussen
Aupal Mondal
Lisa Macias
Chandra Bhat

TxDOT Project 0-7054: Integration of Stated Preference and Revealed Preference Methods in Regional Travel Survey Programs

July 2022; Published October 2022

Performing Organization:	**Sponsoring Organization:**
Center for Transportation Research	Texas Department of Transportation
The University of Texas at Austin	Research and Technology Implementation Office
3925 W. Braker Lane, 4th Floor	125 E 11th St.
Austin, Texas 78759	Austin, Texas 78701

Performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Table of Contents

1. Introduction

In a survey, revealed preference (RP) questions inquire about a respondent's actual behavior, or behavior that they exhibited in the past. Such questions include how often respondents carry out an activity, what their experience was, or how they would rate said experience. These types of questions, with an adequate sample size of respondents, can provide researchers useful information about the current climate of any transportation issue. A downside of RP surveys is that they are unable to inquire about the impacts of environments that do not yet exist. For these types of scenarios, a different type of question is used—referred to as a stated preference (SP) question, which ask about choices that respondents *would* make in a hypothetical scenario.

This guidebook will outline the design, development, and deployment of an SP experiment alongside an entire RP-SP survey to best harness the information provided by respondents on behaviors, attitudes, and reactions to both the future and past, the hypothetical and real. Once the RP-SP survey has been designed and deployed, responses have been collected, and a modeling process completed, the final step will be to determine how the data can best supplement the current analysis method used in traditional travel demand modeling strategies by TxDOT. The results of an effectively designed and deployed RP-SP survey may point to a new direction for potential investigation. This guidebook will explain the RP-SP survey development strategy using the following seven steps:

1. Scenario Development and Risk Assessment/Reduction
2. Survey Content Design
3. Survey Administration
4. Deployment Strategy and Data Collection
5. Data Assembly
6. Data Analysis
7. Integration of Results and Data

The following sections describe a procedure for fully developing, deploying, analyzing, and integrating RP and SP methods— more specifically, integrating SP experiments into RP surveys. From this guidebook, an RP-SP survey may be developed on any topic of interest. The Workplace Location (WPL) RP-SP survey and its design, deployment, and analysis processes, will be referenced as an example throughout the guidebook.

2. Fundamental Techniques for RP-SP Survey Creation

This guidebook identifies some key techniques, based on advantages and limitations of past SP surveys, underlying successful RP-SP survey deployment; these will be discussed in more detail in the next section's step-by-step process for developing an SP survey:

- *use logic methods* to create linkages between RP and SP questions when presenting SP experiments to respondents, based on preceding RP responses, as this ensures realistic scenarios are presented to each individual and limits uncertainty in later analysis;

- *jointly model RP and SP data* to harness the advantages (and compensate for the disadvantages) of each type of data;
- *frame the hypothetical scenario* with specific intent and clearly define each term to obtain reasonable and consistent responses; and
- *define each attribute and attribute level* included in the SP experiment to clearly and uniformly convey the idea behind each component.

3. Step-by-Step Guide to Developing an RP-SP Survey

3.1 Scenario Development and Risk Assessment/Reduction

3.1.1 Determine the topic of interest

Begin with determining the topic of interest for the SP experiment and the best hypothetical scenario to gather data on this topic to inform decision-making. The results of the survey data collection and analysis efforts should lead to insightful conclusions that are relevant to policy actions. The goal is to inform decision-making, rather than simply collect data to serve as a basis for statistical modeling. Additionally, the results should lead to researchers' exploration of what the best kind of travel model to develop from the additional information obtained through the SP questions. Therefore, the framing and content of the SP questions must be relatively specific. Typically, when researchers design surveys for transportation-related purposes, SP experiments are used to gather topical information on three general themes:

- Anticipating the effects of the accelerating pace of transportation technology development
- Determining the impact of complex government transportation (and other) policies
- Establishing a use case for large-scale infrastructure projects

More recently, a fourth theme has become increasingly imperative to study:

- Assessing the impact of the COVID-19 pandemic on future travel behavior and transportation networks

However, SP questions are not limited to these four themes. Common hypothetical scenario categories include:

- autonomous vehicles
- congestion pricing
- managed lanes
- facility improvement
- freight transportation
- public transit/mode choice
- Impact of COVID-19 on Travel Behavior
    - teleworking
    - school

- air travel
- e-commerce
- public transit/mode choice
- transportation network companies (TNCs)
- long-distance travel

Overall, hypothetical scenarios can be developed about any other topic related to transportation planning.

3.1.2 Assess the risk and uncertainty of the scenario

Once the topic of interest has been determined, the risk and uncertainty of the scenario should be assessed. The hypothetical nature of any proposed SP scenario is accompanied by a certain level of uncertainty in the topic and context of the study. This uncertainty arises from two main sources:

1. Study Design: whether the study combines RP-SP questions or employs only SP questions in its design and analysis without developing a clear context and description of the hypothetical scenario.

2. Topic and Context of Study: whether the hypothetical scenario exists in other regions but not in the area of interest to the survey (such as congestion pricing), or if the scenario involves advancing technology or other situations unfamiliar to most individuals, such as autonomous vehicles (AVs).

The combination of these two issues qualitatively defines the risk associated with each SP survey for the hypothetical scenarios in consideration. Additionally, the risk level inherent to each of the hypothetical scenario categories will significantly influence the overall risk associated with the data collected from a survey. The degree to which respondents can envision and are willing to adopt the proposed service or facility influences the level of risk associated with the survey results: the less familiar or prevalent the scenario, the less reliable the answers, resulting in more uncertainty and potential risk for the entity conducting the survey. In order to limit risk and instill increased confidence in decisions and investments based on the SP data analysis, researchers should write scenario descriptions that focus on topics and contexts more familiar to their proposed sample population. However, this is not always possible, as many scenarios, such as regular use of fully autonomous vehicles (AVs), are relevant to a situation that has no parallel in today's travel environment. This makes it difficult for transportation engineers and planners to develop questions and analyze responses, and for respondents to answer questions about a reality they must conjure up based on hypothetical AV technology scenarios.

Therefore, when developing the SP scenario, acknowledgement of the uncertainty associated with it is necessary. Based on the topic and context of the study, the potential risk of relying on that study's results can be assessed and assigned to one of three categories. It is essential to note that this three-level categorization of risk is a subjective and qualitative determination based on previous experience with survey design and response modeling. Following is a description and example of each risk level:

- *Low confidence*
    - This survey posits unfamiliar scenarios and asks only SP questions (without clearly positioning the hypothetical scenario), and thus presents a high level of risk to anyone relying on the responses; the modeler needs to be cautious in their analysis of the data.
    - Example: a survey assessing potential usage of AVs, which can be presented only in an SP format, because the majority of respondents have never experienced or interacted with AVs on a regular basis and on standard roadways.
- *Medium confidence*
    - This survey presents more familiar scenarios, using both RP and SP questions, or using only SP questions with only an adequate description of the context of the SP hypothetical scenario.
    - Example: if congestion pricing is being considered for a currently untolled roadway and the survey contains RP-SP formatted questions, the modeler can be moderately confident in the respondents' ability to predict accurately their responses to the scenarios presented, and thus can have more confidence in the resulting data analysis (as compared to topics falling into the low-confidence category). However, while respondents already know the facility and are familiar with the concept of congestion pricing, that particular form of tolling is not already implemented on this specific roadway. Thus, this scenario does present some potential risk, and the survey results cannot be analyzed with a high level of confidence.
- *High confidence*
    - This survey references a well-established technology, practice, or facility, using both RP and SP questions, or a clearly described hypothetical scenario in a case with only SP questions.
    - Example: a survey to determine the impacts of raising existing parking fees. Using both a familiar context and a combined RP-SP survey creates a high confidence in the results and analysis. Parking fees already exist, and respondents are just being asked to decide between either paying an increased fee or parking elsewhere. There is high confidence that the responses will align with the actual choices individuals would make in this situation, and therefore low potential risk in the question.

Of course, not all scenarios may fit well within this rather aggregate risk level classification. In some contexts, using only SP questions (without the combination of RP question) pose less risk and higher confidence in the responses. For example, consider work place location choice in a future where the effect of COVID on health considerations has waned, though it may still be a factor in decision-making, especially for immunocompromised and other specific segments of the population. In this situation, the act of working from home or from the work place (and the very act of working) is something that respondents can fathom and understand (therefore the

hypothetical scenario is clearly defined). In this situation, after a large shock such as COVID, it may make little sense to combine RP-SP data in the context of work place location choice before COVID (which would be collected through RP questions) and what an individual may do in a future where COVID becomes endemic (which would be collected through SP questions), because so much has changed in respondents' lives due to the pandemic. Thus, combining RP-SP questions would actually be at the lowest end of the confidence scale in this context, while using SP choice alone (with carefully worded and designed SP questions) can present a lower risk, given the familiarity of the activity (work) and the work place locations (home or regular work place). Thus, the context of the study matters too.

3.1.3 Reduce and mitigate the risk and uncertainty

After thoroughly assessing the risk, it is important to develop strategies to best reduce and mitigate the risk and uncertainty in the SP survey. This involves the recognition and premeditation for the next step, *Survey Content Design*, when all components must be carefully and cohesively constructed and integrated.

Strategies include, but are not limited to:

- clearly and uniformly conveying the idea behind the hypothetical scenario to all survey respondents by carefully crafting the wording and descriptions to ensure that the respondent comprehends the question presented in the context of the specific scenario envisioned by the researcher;
- anchoring the SP experiment to related and carefully crafted RP questions, so that both can be used alongside each other later in the analysis process;
- linking the SP experiment to RP questions so that specific alternatives, attributes, attribute levels, or the entire SP experiment are shown only to respondents to whom they are applicable, in order to make the experiment relatable to each respondent; and
- designing the attributes and their levels so that their values or circumstances are backed by prior research and are applicable/realistic to the region of study in order to ensure the most accurate scenario for the respondents.

Each of these strategies will be discussed in more detail in later steps, but it is important to recognize different strategies' potential for reducing risk by increasing realism and consistency within and across the hypothetical scenario for all respondents.

3.2 Survey Content Design

3.2.1 Determine the SP elicitation mechanism

Now that the topic of the hypothetical scenario has been determined, the details of the experiment must be designed. This starts with deciding on the survey format, also referred to as the *SP elicitation mechanism*. Travel surveys frequently use the following elicitation mechanisms:

- Contingent Behavior
    - This format asks the respondent what they would do in a hypothetical scenario. These questions do not have varying attribute levels and instead ask a respondent to answer what they would do **if** a situation occurred.
- Contingent Valuation
    - This format asks a respondent to consider the value that an option holds for them. They may be asked whether or not they would choose an option given its value, or how much they are willing to pay for an option. The respondent will see a set of these SP questions, with varying attribute levels across the set; these attribute levels are presented either in the question or in the response options. The varying attribute levels would be the value or the cost of the option.
- Choice Experiment
    - In this format, respondents are presented with a set of SP questions and instructed to choose from, or rank, two or more alternatives with varying attribute levels across the SP question set.

Both contingent behavior and contingent valuation mechanisms are primarily used to obtain more general behavior intentions in a hypothetical context. This helps to forecast broad future trends for more general policy insight. Contingent behavior SP questions are easier for a respondent to understand, as the hypothetical scenario presented to them appears less complex. Contingent valuation SP questions help to put a price or monetary value on a specific change to a system, activity, or commodity. A choice experiment, on the other hand, offers a multidimensional hypothetical scenario, with exact values and conditions, that allows modelers to analyze specific behavioral responses for quantitative statistical models. Though a choice experiment with multiple questions, alternatives, and attribute levels may seem a bit overwhelming to a respondent, the data gathered from such a set of SP questions is extremely valuable for projecting a wide range of future travel behaviors and demands. Overall, each SP question format offers an efficient method to collect responses on predicted future behavior in a situation that may be unknown, such as the evolution of lifestyles and attitudes in a post-COVID world.

The contingent behavior and contingent valuation mechanisms are fairly straightforward to design and integrate into an RP-SP survey. Therefore, the remainder of this step (and guidebook) will focus on designing, deploying, and analyzing a survey that contains a choice experiment.

3.2.2 Design the SP experiment

Once the best SP elicitation mechanism has been determined for conveying the hypothetical scenario and future analysis opportunities, it is time to design the actual experiment. Some components of the SP questions must be crafted with particular care.

3.2.2.1 Pose the general question

To begin, a general question must be posed to guide the rest of the experiment's design. This question will be a more specific instance of the chosen topic and will aim to hone in on the purpose behind the survey. For example, for a survey on [chosen topic], the general question might be "as individuals emerge from the COVID-19 pandemic, what will be the ongoing travel impact of the shifts in work-from-home behavior and commuting patterns?"

3.2.2.2 Plan the alternatives

Once the general question is put in place, plan the alternatives. To reiterate, the alternatives are the response options offered to the respondent to consider after the specific scenario has been presented in full (i.e., the context or framing and the attributes and their levels).

For example, when presenting a hypothetical scenario about ideal workplace locations, the alternatives may include: work from home, work from the in-person work office, and work from a third workplace.

3.2.2.3 Design the attributes and their respective levels

The alternatives act as a guide for the next task, which is the design of the attributes, as well as their levels. The attributes are used as variables within the experiment, so that different scenarios can be presented to respondents, though it is important that the attribute levels come across as realistic. Then, based on these variables, respondents will select the option among the presented alternatives they would be most likely to choose if they were in the situation.

For example, when in the same ideal workplace location scenario, attributes may include workplace distraction levels or COVID threat level, where the attribute levels for these two attributes may be none, low, or high.

It is important that every attribute is qualified or quantified by varying levels that are realistic, whether they are backed by data about a specific region or are simply important for future analysis purposes. In each experiment shown to each respondent, that respondent will see only one level for each attribute in the presentation of the scenario or in each of the alternatives. The combination of the attribute levels across all attributes (which will essentially define, frame and allow for variation across the contexts of the different hypothetical scenarios) will be decided (and optimized for analysis purposes) later through the experimental design process.

It is also important that the survey design contains neither too few nor too many attributes or attribute levels. Typical SP experiments have four or more attributes. There are usually two to six levels for each attribute. With too few attributes or levels, variety across all scenarios is limited for respondents. With too many attributes, the hypothetical scenario gets too complicated to imagine. If the number of attribute levels is excessive, the possible variations of the experiments will increase exponentially, which will complicate later data analysis procedures and require a larger sample size to obtain an adequate number of responses for each SP question.

3.2.2.4 Frame the SP experiment

Next, after the alternatives and attributes have been designed, it is time to cohesively assemble the question. This begins with the transformation of the general question into an official and well-thought-out framing of the SP experiment. Within an SP survey, specific introductions to each question set the scene for the respondent, framing the hypothetical scenario with the intent to obtain reasonable responses. It is critical that every respondent understand the framing of an SP question. Given the hypothetical nature of the scenarios posed in SP surveys, a level of uncertainty accompanies the question, as previously addressed. Therefore, it is vital that the survey designer clearly and uniformly conveys the idea behind the hypothetical scenario to all survey respondents. Methods used to frame the experiment include:

- Defining the technology or status quo in question. For example,
    - "An *Autonomous Vehicle (AV)* is a vehicle that drives itself without human supervision or control. It picks up and drops off passengers including those who do not drive (e.g., children, the elderly), parks itself, and picks up and delivers laundry, groceries, or food orders on its own. When answering the questions in this section, please assume a future in which AVs are widely adopted, but human-driven vehicles are still present."
    - "Imagine COVID-19 continues to impact our everyday lives, however 90% of the population has been vaccinated and the disease's spread is under control."
- Providing a detailed description of each alternative. For example,
    - "*Working from home* indicates that you work every day from your place of residence and do not have a daily commute outside of your home."
    - "*Working from (in-person) work office* indicates that you have a daily commute to your place of work."
    - "*Work from a third WPL* includes locations like a coffee shop, a designated co-working space, a hotel, or a restaurant but does not include working from a client's site, which would instead be categorized as the Work Office. The appeals of a third workplace may include less distractions than when working from home and not having to commute all the way to the regular workplace. However, a third workplace may still have crowding and require a commute."
- Providing a detailed description of each attribute, as well as each attribute level. For example,
    - "*Level of distraction at the workplace*: How crowded your workplace is and the distance between you and other people and their workspaces. The options: the outside-of-home workplace is crowded and you are in close proximity to loud coworkers (*High distraction*); the outside-of-home workplace is crowded and you are in close proximity to quiet coworkers (*Low distraction*); there is some crowding at the outside-of-home workplace, but you have a small area to yourself or with chosen coworkers (*Extremely low distraction*); and no crowding at the outside-of-

home workplace, and you have your own designated, quiet, closed-off room *(No distractions)"*

- "*Workplace safety implementation for COVID*: How your workplace is addressing COVID-19 and implementing safety precautions and regulations. These may include requiring face coverings, implementing social distancing, introducing hand sanitation stations, placing barriers between workspaces, or mandating COVID-19 testing and/or vaccinations."

3.2.3 Determine and implement the experimental design

Once the attributes and their levels have been determined, the possible combinations of attribute levels shown to respondents must be designed. To best harness the range of information SP questions can gather, their experimental design is important. The SP component must be designed in such a way that information on a wide range of possibilities within the scenario of interest can be obtained using a minimal number of questions. This is a critical issue, since the burden on respondents of increasing the length and complexity must be carefully evaluated and balanced against additional information that can be elicited through SP questions. Since it may not be possible to ask every possible version of the SP question to enough respondents, mathematical design methods for experimental survey design can be employed to select an optimal subset of instances that would be most useful for predicting travel behavior. These various experimental survey design methods are used to manage trade-offs to maximize the success of a study. It is important to note that experimental design is not necessary for contingent-behavior-type questions, but it is imperative for the contingent valuation and choice experiment SP formats. Table 1 provides a brief overview of four statistical experimental design types typically used in SP surveys. Intricacies of the benefits and workings of each experimental design are not important for a survey designer to know. Any of the four designs will be adequate for developing a set of SP questions. This overview is presented to inform the designer of the existing experimental design options, rather than teach them how to use them, as there are automated software that perform the process for the designer.

Table 1: Statistical Experimental Designs

Type of Experiment	Characteristics
Full Factorial Design	Each level of each attribute is combined with every other level of every other attribute. For example, a design with two attributes with three levels each and two attributes with two levels each could have 36 scenarios or subsets ($3^2 * 2^2 = 36$). This design captures all the main effects and interaction effects of attributes within the dataset.
Fractional Factorial Design	When not all interaction effects are equally important, some can be ignored. This type of experiment design allows for the reduction of a large volume of scenarios created by the full factorial design by ignoring some interactions of attributes.

Type of Experiment	Characteristics
Orthogonal Design	All attributes are statistically independent of one another. Main effects are the focus of study and analysis in such a design
Efficient/Optimal Design	This method optimizes the amount of information obtained from a design, accomplished through multiple methods (such as D-efficient design).

An optimal subset of SP questions is selected for each individual to respond to through the experimental design process. Each statistical experimental design type included in Table 1 can be automated through various experimental design software programs, such as the software SPSS. Alternatives, attributes, attribute levels, and desired number of questions or scenarios for each SP subset are input into the experiment design software by the survey designer. The program will consider all possible combinations of attribute levels for each alternative and question and run the preferred experimental design algorithm, outputting an optimal subset for the designer to include in their survey. The array of selected questions highly depends on which experimental design type is chosen; orthogonal design is most commonly used, while D-efficient is a close second. Each experimental design type functions as a premeditated method to randomize the selection of attribute levels to be included in a set of SP questions.

The experimental design should be conducted such that the SP data can be optimally combined with RP data to extract the most information possible for modeling structures. When determining the attribute levels for varying alternatives, it is important to ensure realism in the choice situation. This is done by making sure the attribute values are realistic to the respondent, so they are able to imagine the alternatives with some existing memory and reference. For example, if most respondents report commuting times around 30 minutes, a level of two hours for the travel time attribute would create an unrealistic scenario for the respondent and increase uncertainty in the accuracy of their response. It would also limit the modeling possibilities of linking current commute patterns (RP data) with future hypothetical commute patterns (SP data), because the RP and SP data impose such different travel times that they would almost be incomparable. This "unrealistic" attribute level should not be presented to any of the respondents and can be simply removed from the set.

It is also important to choose how many SP experiments each respondent will be presented with. Typically, asking two to three questions of each respondent balances reducing respondent fatigue and gathering enough data for the chosen scenarios. When coding the SP experiment into the survey software, the modeler can simply set the logic such that each respondent is presented, at random, with a fixed number of SP experiments resulting from the experimental design process.

3.2.4 Employ RP questions

Several RP questions are asked before and after the SP portion of the experiment to qualify respondents for participation. Determine which RP questions to employ, as these questions must be curated with a specific analysis intention in mind.

a. Which RP Questions to Employ

- Choose which RP questions are imperative to the analysis of the desired SP questions. In order to obtain necessary data, choose and form these questions carefully, even if their only role is to "frame" the SP questions.
    - The position and placement of RP questions within the survey is up to the discretion of the survey creator. Typically, several RP questions are asked before and after the SP portion of the experiment to qualify respondents for participation.
    - In some circumstances, placing RP questions prior to SP questions allows the survey developer to lead the participant through their survey experience. By arranging the questions to improve flow, the developer can ease survey participation fatigue.
- Suggested RP Questions
    - Individual-Level Questions
        - Answers to sociodemographic questions help modelers account for heterogeneity in preferences among respondents. These questions can include:
            - Gender
            - Age
            - Income
            - Employment type
            - Education level
            - Residence in Texas
            - Driver's license status
    - Household-Level Questions
        - These questions gather more information on a respondent's household environment to offer additional insight on heterogeneity of respondent preferences. These questions can include:
            - Household income
            - Household size
            - Presence of children
            - Number of vehicles

3.2.5 Set up linkages between the RP and SP components

A next step in the survey design is setting up linkages between the RP and SP components so the experiment, or specific alternatives, attributes, or attribute levels, are only revealed to qualifying respondents. Once the experimental design is configured and the set of SP questions is determined, the survey designer must decide which (if not all) respondents will see and answer the set of SP questions (also known as the SP component) as they proceed through the survey. To some respondents, the SP component will not be applicable. For example, a respondent who lives just a couple of blocks from work (as revealed by RP questions answered earlier in the survey) should not be presented with a hypothetical situation where they have to choose between rail and transit to go to work after the pandemic. The SP component should be anchored to the RP component so that the hypothetical situation developed in the SP component is relatable to the specific survey respondent. This, once again, helps provide some realism to respondents, so that they can better place themselves in the hypothetical context.

Online survey administration programs can use a respondent's answer to an earlier RP question to determine the attributes characterizing the SP experiment they are presented with. Most online tools allow for multiple RP questions to be linked to a single SP component, making it easy to construct appropriate SP scenarios. If this RP-SP linkage in constructing SP scenarios is forgone, it would increase the uncertainty level in the SP data collected and reduce the validity of behavioral projections. By linking RP questions to the SP scenario presented, survey developers can often increase the relevance of the question to the respondent, thereby increasing certainty in the data and the validity of projections. However, some SP questions may not benefit from linkage to RP questions, as the topic is universally applicable to all respondents chosen to participate in the survey. For example, if a survey is being presented only to a group of individuals who have previously reported that they are employed in a given region, an SP component presenting alternative routes to work with varying travel times and toll costs will be applicable to all respondents taking that survey.

3.2.6 Conduct a friends-and-family pilot of the survey

Before the SP experiment and the entire survey is finalized, it is helpful to initiate a few rounds of a friends-and-family pilot of the survey. The goal of the pilot survey is to ensure that each question makes sense to respondents and is presented in a streamlined fashion. Before it is deployed to a broader public, the pilot survey is deployed to a small sample, typically of friends, family members, and coworkers, with the following goals:

- Obtain general feedback
- Clarify confusing or ambiguous terminology
- Test the flow and logic of the entire set of RP and SP questions
- Confirm all combinations of attribute levels within each SP experiment are realistic
- Address other potential response issues

3.3 Survey Administration

During the process of designing both the SP experiment and the entire RP-SP survey, it is important to design for the administration method that will be used. The design considerations are specific to the survey medium. For example, an online or web-based survey conducted through a multimedia device (such as a tablet or phone-based app) would differ in design from one that is conducted by phone. Other survey administration options include:

- In-person at respondent's home
- In-person at a set location
- In-person on or outside public transportation
- Over the phone
- Mail-in response

Today, the majority of surveys are administered on an online platform, allowing respondents to use their smartphone devices or computers for answering from their home, work, or on the go. However, some populations, such as older generations or lower-income households, are less responsive to online surveys, and other capture mechanisms are needed to gather their responses. Though an SP survey could hypothetically be administered either over the phone or in person, the complicated and hypothetical nature of the design of an SP experiment and its variable attributes mean an online platform is far more suitable, allowing for easy coding schemes and implementation of logic structures within the survey. In an online platform, the attribute values and responses are automatically recorded in digitized form, making this a very convenient SP survey administration approach. If the survey is administered over the phone or in person, it will be up to the survey administrator to determine and keep track of which SP questions from the complete set are asked. Specific attribute levels must be noted when the responses are coded into an aggregate dataset in order to perform proper modeling. Once again, an online platform, such as Qualtrics, allows for ease in design, deployment, and post data processing.

3.4 Deployment Strategy and Data Collection

3.4.1 Consider the survey deployment method

Before the survey is deployed, a few strategic decisions must be made. To begin, the survey deployment method must be considered. Developing a strategy to best elicit responses, in the most cost-effective manner, is vital in order to ensure a broad range of respondent participation. In some instances, incentives can be offered to respondents to increase the likelihood of their completing the survey. Deployment should not be limited to only one strategy; multiple strategies can be used so that responses from the most relevant/representative populations responses are collected. Options include, but are not limited to:

- Social media outlets
- Hiring consultants/experts to handle deployment
- Professional networks

- Distribution of postcards with an online link
- Phone surveys
- In-person surveys on public transportation
- Door-to-door surveys

3.4.2 Set a desired sample size

Before the survey is deployed, a few goals must be set regarding responses. First, determine a desired sample size and devise appropriate strategies to achieve it. Goals regarding representation will be covered in the next section. Sample size determinations are quite straightforward for univariate statistics such as estimating the mean income in the population. However, there are no clear theoretical formulas for most choice models. This is because there are multiple considerations, including the number of exogenous variables, the functional form of the effects of variables, interactions of exogenous variables, the range (variance) of the exogenous variables, and, in the case of choice models, the number of alternatives and the split share of the alternatives. In the case of RP-SP studies, the number of attributes and their levels, the number of SP questions, the likely parameter estimates, and again the SP experiment alternative(s) that individuals may choose, are all considerations for setting the desired sample size. Many of these will not be known in advance, as the SP portion is an actual experiment, with inherently unknown results. Even so, there is no clear theoretical formula to estimate sample size in multivariate RP-SP models, as there is for univariate descriptive statistics.

In typical RP-SP data collection and experimental designs, a sample size of 300 to 500 is considered the minimum. However, even a sample of this size may not be adequate to ensure enough respondents choose each alternative (this would happen if the expressed choices are heavily skewed toward one or more alternatives, with a very low share of one or more other alternatives; this can be controlled, to some extent, by developing attribute levels appropriately based on current RP choices, but there is still limited control here because SP experiments concern theoretical or future situations). Furthermore, in choice models, not only do the overall alternative shares across all respondents matter, but so do the number of individuals choosing each alternative within each demographic segment of potential relevance. For example, if gender may impact the dependent variable, it is important that, within each group of males and females, at least 50 to 100 individuals are choosing each alternative, or else most choice model will inaccurately estimate the effect of being a female. Thus, it is generally prudent to go well beyond a sample size of 300 to 500. Once representation considerations are addressed, however, there are diminishing marginal returns from increasing sample size.

The general consensus is that a sample size of 1,100 to 1,200 respondents is ideal for RP-SP analysis, though most survey designers prefer to achieve approximately 1,400 to 1,500 respondents.

3.4.3 Make decisions about gathering a representative sample

The second goal concerns gathering representative samples. While a complete representation of the region under study is not required for most modeling purposes, it is important to set some standards about who must or must not be under- or over-represented in a sample. While estimated behavioral relationships among variables would not necessarily be affected by a non-representative (based on demographics) sample, application of estimated models to examine effects of a specific policy would be affected. To be usable for the evaluation of the impacts of interventions, policies, or future projections, the sample may have to be weighted post-data-collection to be representative of target residents of the region being studied. However, for any model estimation, an adequate number of observations representing each potential subgroup of importance is necessary. As an example, for modelers to identify a gender identity–based difference (after controlling for other variables) in AV adoption, the sample must include adequate numbers of individuals who identify as men and as women to capture the effect. Researchers can accomplish this by targeting groups specifically when distributing the survey, or monitoring data collection so that certain demographics are represented in the compiled and growing dataset (and if a certain demographic is not represented, the researchers will need re-strategize how to target that group mid-data collection). Groups that may need to be encouraged or limited (in no specific order) include older populations, students, employed populations, and minority groups.

3.4.4 Deploy the survey and collect the data

After the survey has been finalized and deployment strategies and goals developed, it is time to put them in motion and deploy the survey and collect data. Once deployment and participant interaction have begun, the survey team must monitor survey activity. The growing dataset from survey participation may include incomplete responses that must not be counted towards the final sample size and, therefore, must be removed from the dataset. For the collected data to be truly representative of predicted behaviors, it must only include data from respondents who remained engaged with the survey from start to finish.

3.5 Data Assembly

Once enough responses are collected (as predetermined during sample size considerations), it is time to organize the data so that it can be efficiently used for analysis purposes within a modeling framework. Typically, an RP-SP dataset consists of sociodemographic and/or household responses, a set of RP responses that reveal current travel behavior, and a set of responses to SP experiments that proposed hypothetical choice scenarios. Therefore, the resulting dataset will contain a significant amount of information, which demands a thorough organization.

3.5.1 Organize RP data

Organizing RP data from an RP-SP survey is fairly straightforward, as it is identical to the organization process for RP data found in a traditional household survey dataset.

If the survey was collected through an online platform[1], the original dataset as downloaded from the site used for distribution is adequate for descriptive analysis, though it will not be in a format that is suitable for modeling. Gender responses, for example, are typically already categorized and organized straightforwardly in the downloaded dataset. By running a simple frequency table, the analyst will be able to determine the number of individuals in their sample who are male, female, non-binary, or other.

For use in regression or other analyses, as well as for general ease of use, the RP data will need to be organized in one of several alternative ways, depending on the nature of the question. Typically, there are four main different types of question and response formats. They are reviewed below, accompanied by a brief R code snippet that provides an example of how the data organization process was performed for the WPL dataset:

1) Binary variables (example: Do you have a driver's license? (Q16)). A respondent answers *yes* or *no* to a question, and their response is converted into a single binary column, displaying 1 for yes or 0 for no. The following code performs this conversion:

```
my_data$license = as.numeric((my_data[["Q16"]] == "Yes"))
```

2) Verbal categories (example: What type of region is your residence located in: urban, suburban, or rural? (Q11)). For this type of question, a binary column must be created for each response option. The following code will convert, for example, the answer *urban* into a 1 (yes) in the urban column and a 0 (no) in the suburban and rural columns.

```
my_data$rural = as.numeric((my_data[['Q11']] == "Rural"))
my_data$suburban = as.numeric((my_data[['Q11']] == "Suburban"))
my_data$urban = as.numeric((my_data[['Q11']] == "Urban"))
```

3) Numerical categories (example: what year were you born in? (Q116)). These numerical responses may either be a) continuous (such as birth year, when respondents can input any reasonable year) or b) already grouped to a certain extent (such as for current commute time, where respondents are asked to select the most appropriate 5-minute increment (Q41.1_3)). It is not uncommon for the analyst to have to regroup elementary categories into broader categories if there are too few responses in any elementary category for a model to distinguish the characteristics of that category from those of others; combining responses into, for example, 15-year age groupings or 20-minute commute increments may improve and simplify the modeling and interpretation. The analyst can test successively broader groupings during the analysis process. Example code for each of these situations is below.

[1] If the deployment process was not limited to an online platform, identical measures to those for RP-only surveys should be taken to fuse and organize responses.

a) **Continuous numbers:** the following code calculates age from birth year (see row 2) and then assigns the respondent to one of eight age groups (see row 3):

```
my_data$age1 = as.numeric(as.character(my_data$Q116))
my_data$age2 = 2021 - my_data$age1)
my_data$age3 = cut(my_data$age1,
                   breaks=c(Inf, 1942, 1957, 1972, 1982,
                   1992, 1997, 2003, -Inf),
                   labels=c("80+", "65-79", "50-64", "40-
                   49", "30-39", "25-29", "18-24", "under
                   18"))
```

b) **Already grouped numbers:** The survey asked respondents to indicate their commute time to their Work Office in the closest 5-minute increment up to 75 minutes, meaning there were 16 possible selections. However, upon review of the number of responses in each commuter time category, it became clear that the analysis could not support such a disaggregate categorization. So, in addition to testing a continuous value of commute time (by ascribing mid-point values for each category, i.e., 3 minutes to the 0–5 minutes category and 90 minutes to the 75 minutes or longer category), the research team also tested a broader categorization of commute time (note that doing so allows a potentially non-linear effect of commute time on the decision variable of interest). The following code groups commute times into seven different increments:[2]

```
my_data$comm0 = as.numeric((my_data$Q41.1_3 == 0))
my_data$comm10 = as.numeric((my_data$Q41.1_3 <= 10) &
(my_data$Q41.1_3 != 0))
my_data$comm25 = as.numeric((my_data$Q41.1_3 > 10) &
(my_data$Q41.1_3 <= 25))
my_data$comm45 = as.numeric((my_data$Q41.1_3 > 25) &
(my_data$Q41.1_3 <= 45))
my_data$comm60 = as.numeric((my_data$Q41.1_3 > 45) &
(my_data$Q41.1_3 <= 60))
my_data$comm75 = as.numeric((my_data$Q41.1_3 > 60) &
(my_data$Q41.1_3 <= 75))
```

4) Likert scale groupings (example: How satisfied are you with your current commute to your Work Office: extremely dissatisfied, somewhat dissatisfied, neither satisfied nor dissatisfied, somewhat satisfied, or extremely satisfied? (Q49)). The analyst will assign an ascending value to each of the Likert measures, with the most negative response (which may be in the form of "dissatisfied" or "strongly disagree") assigned a value of 1, the next most negative ("somewhat dissatisfied" or " somewhat disagree") assigned a value of 2, and so on. There may be three, five, or even more categories of the Likert scale, with the highest value equaling the number of possible responses. A neutral option ("neither

[2] "==" means "equal to," "<=" means "equal to or less than," ">=" means "equal to or greater than," and "!=" means "does not include."

satisfied nor dissatisfied" or "neither agree nor disagree") is typically included, which will fall in the middle of the answers and should then receive the middle, or median, value (regarding commute time, "neither satisfied nor dissatisfied" was assigned a value of 3).

```
my_data$comSat = car::recode(my_data$Q49,
                   "'Extremely dissatisfied' = 1;
                    'Somewhat dissatisfied'  = 2;
                    'Neither satisfied nor dissatisfied'  = 3;
                    'Somewhat satisfied'  = 4;
                    'Extremely satisfied' = 5")
my_data$comSat = as.numeric(my_data$comSat)
```

3.5.2 Organize SP data

The SP experiment data typically requires a considerably greater amount of organization compared to the RP data. As previously reviewed, each respondent is typically shown two to four different SP questions with varying attribute levels (a single SP question will be referred to as a *choice occasion* for the remainder of this discussion), alongside an array of RP and sociodemographic questions. To keep analysis efforts as simple and straightforward as possible, the SP dataset should be organized so an individual's response to each choice occasion (along with the occasion's attribute level values) becomes its own row in the dataset. Each row should also contain that individual's sociodemographic and household characteristic data as well as their current travel behaviors extracted from the RP questions (or RP choice occasions).[3] Thus, some repetition occurs within the dataset structure, as the same demographic and RP information will appear on both rows holding an individual's responses to each SP choice occasion they were presented with.

The manual duplication process can be performed easily in Excel. Automating it should be possible through R as well. Figure 1 provides a visual example of this process in which two SP questions were presented to each respondent. First, open the dataset in Excel as a .csv file (Step 1 shows all the data while Step 2 demonstrates that the RP data will remain unchanged). Next, copy all of the rows of respondent answers (i.e., all except the top row of headers) and paste them below the last row of responses (Step 3). Highlighting either the original or duplicated responses can create a useful visual indicator of the delineation. From the first set of responses only (shaded gray in Figure 1), delete the contents of all cells related to Scenario B (Step 4). Then, from the second set of responses (shaded yellow in Figure 1), delete the contents of all cells related to scenario A (Step 4). Cut the contents of the Scenario B–related columns in the second set of responses and paste them into the now-empty Scenario A columns (Step 5). The new dataset contains each respondent's SP choice occasion data in a separate row, alongside their RP data, and the highlighting used to distinguish the sets of responses can now be removed (Step 6).

[2]This is effectively treating each SP choice occasion as repeated choice events from the same individual.

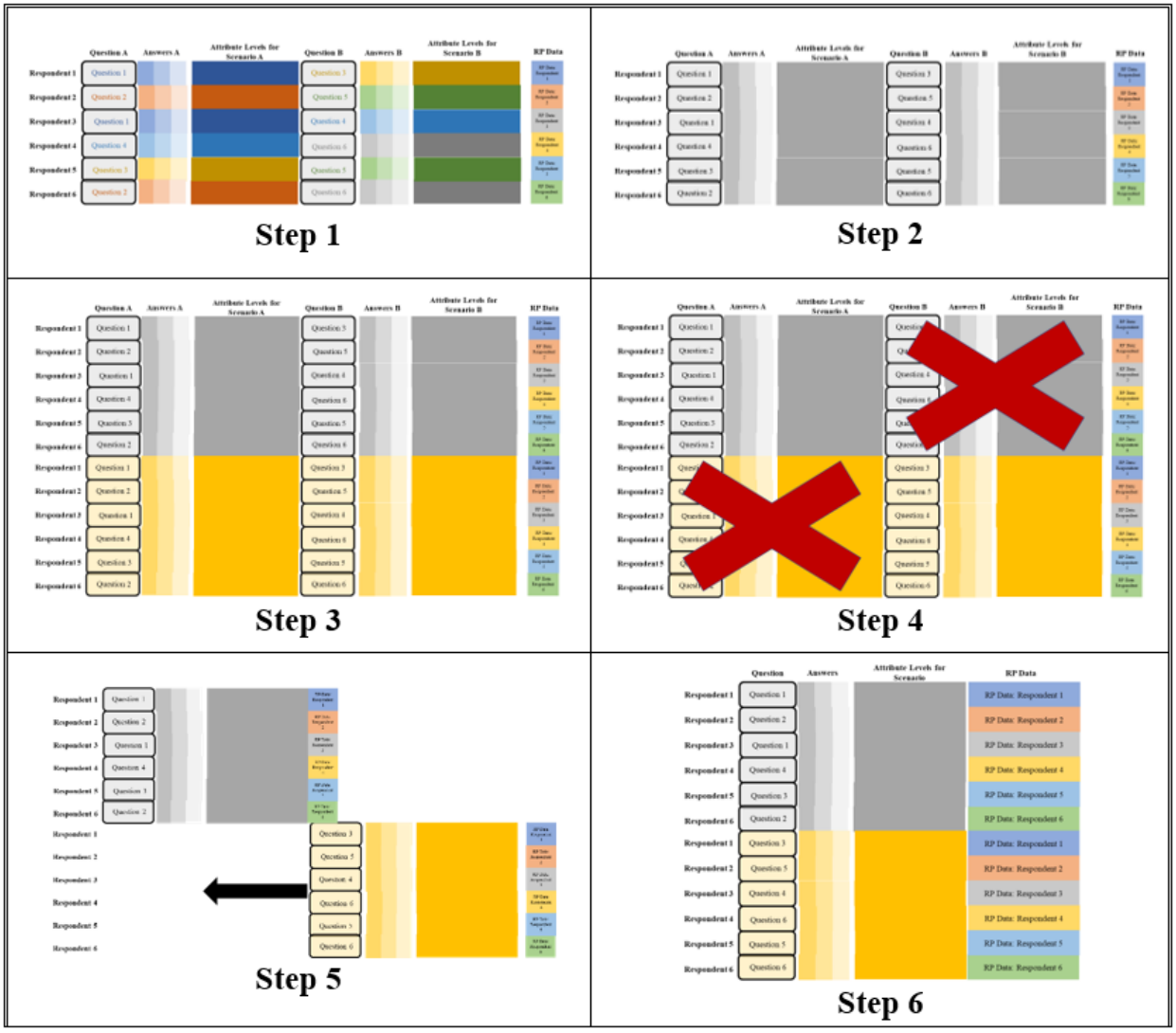

Figure 1: Visualization of the duplication process

A visualization of an organized database is provided in Figure 2 (though it excludes most individual and household sociodemographic data for simplicity). Note that each respondent is listed twice in the dataset, with their RP data replicated exactly, though their SP choice occasion, its attribute levels, and their response differs between the two rows. This demonstrates that each respondent was asked two different SP questions, and the dataset now consists of two rows for each respondent. This duplication has no effect on later estimations and provides a straightforward method of organization during the modeling process.

ResponseID	RP Questions				SP Experiment							
	Work Status Before COVID	Work Status Now	Frequency of Telecommuting each week before COVID	Household Size	Scenario	WPL: WFH[12]	WPL: WFO[13]	WPL: WF3[14]	COVID Risk Level	Distraction Level at Home	Crowding at the Workplace	Safety Precautions at Workplace
R_1	Unemployed	Employed	3	2	1	10	4	8	1	3	2	1
R_1	Unemployed	Employed	3	2	37	22	0	0	3	3	3	3
R_2	Unemployed	Employed	2	5 or more	4	4	18	0	2	1	1	2
R_2	Unemployed	Employed	2	5 or more	19	7	12	3	3	1	3	1
R_3	Employed	Employed	3	4	23	15	0	7	1	1	3	2
R_3	Employed	Employed	3	4	32	15	0	7	1	1	1	1

Figure 2: Example of an Organized Dataset for Use in Modeling an SP Experiment

This organization can also be described in mathematical terms. Assume X number of respondents to a survey, with each respondent answering two SP choice occasions. Let there be Y currently observed RP choices, and an array of sociodemographic/household questions. The resulting data may be housed within a dataset that has X rows, each containing two SP choice occasion columns, Y RP choice indicator columns, and multiple columns for the sociodemographic/household information responses. But, for joint RP-SP estimation, it is convenient to translate this dataset structure (with X rows) to a new dataset structure with 2*X rows, each row holding one SP choice occasion, Y RP choice columns (with the RP data in each of the two rows for the same individual having identical entries), and the array of sociodemographic/household columns (again, with this information being identical in the two rows from the same individual, as shown in Figure 2). See Technical Memorandum 11 for a more detailed example on this process.

3.6 Data Analysis

There are multiple ways to analyze RP-SP data, and most of them are identical to RP data analysis approaches and what is done with traditional household survey data before it is used as an input in travel demand models. Two of the main analysis methods that benefit SP data are descriptive statistical analysis and choice modeling.

3.6.1 Descriptive statistical analysis

Employing a descriptive statistical analysis process is an easy but effective way to both get to know the dataset and decipher single- and multivariate trends throughout the data, without requiring extensive modeling.

3.6.1.1 Determine who is in the sample: a sociodemographic analysis of RP data

First, perform a descriptive statistical analysis of respondents' sociodemographics to get to know who is in the dataset. Understanding the sample as revealed by the dataset is helpful in realizing its representativeness of the entire population (specifically if any sociodemographic group that the survey team did not review the sample for was overlooked when monitoring during the deployment process), determining if any variables need to be organized more than was performed through the four strategies listed in Section 3.5.1.1, and developing insight into which sociodemographic variables should be included in further descriptive and choice modeling analysis. It is beneficial to statistically explore the spread of important sociodemographic/household variables that were specifically chosen to include in the survey during the design process, as reviewed in Section 3.2.4. Such RP variables may include:

- Gender
- Age
- Education level
- Household income
- Household structure
- Employment characteristics

- Residential characteristics

These sociodemographics can be analyzed as a univariate through frequency tables, or as a multivariate with crosstabs of different variables. One crosstab it is always recommended to assess is gender and age. It is likely that both of these variables will have a significant impact on the main outcome variables for the choice modeling process. An example of this crosstab from the WPL data can be found in Table 2, which consists of only employed individuals.

Table 2: Sociodemographic Crosstab Example: Gender and Age

Age Group	Gender							
	Female		Male		Non-binary		Total	
	Count	%	Count	%	Count	%	Count	%
18 to 24	10	0.8%	7	0.5%	2	0.2%	19	1.5%
25 to 29	40	3.1%	28	2.2%	1	0.1%	69	5.3%
30 to 34	33	2.6%	31	2.4%	1	0.1%	65	5.0%
35 to 39	53	4.1%	33	2.6%	0	0.0%	86	6.7%
40 to 44	84	6.5%	46	3.6%	0	0.0%	130	10.1%
45 to 49	75	5.8%	41	3.2%	1	0.1%	117	9.1%
50 to 54	138	10.7%	74	5.7%	2	0.2%	214	16.6%
55 to 59	112	8.7%	79	6.1%	1	0.1%	192	14.9%
60 to 64	97	7.5%	82	6.3%	0	0.0%	179	13.9%
65 to 69	70	5.4%	76	5.9%	0	0.0%	146	11.3%
70 to 74	25	1.9%	22	1.7%	0	0.0%	47	3.6%
75 to 79	8	0.6%	12	0.9%	0	0.0%	20	1.5%
80+	3	0.2%	5	0.4%	0	0.0%	8	0.6%
Total	748	57.9%	536	41.5%	8	0.6%	1292	100.0%

Table 2 also provides an example of the descriptive analysis that should be performed in this step. First, a univariate descriptive analysis can be obtained for gender and age on their own. The WPL sample consists of 57.9% female employees, 41.5% male employees, and 0.6% non-binary employees. As for age groups, 6.8% of the sample is employees between the ages of 18 and 29, 21.7% is between 30 and 44, 54.3% is 45 to 64, and 17.1% are 65 or older. Next, a multivariate analysis can be performed, which can pinpoint the largest gender-age group; in the case of the WPL sample, this is women aged 45 to 64 (32.7%), followed by men of the same age group (21.4%). As discussed earlier, it may be helpful to compare these descriptive statistics to the entire employed population of Texas in order to evaluate the representativeness of the sample.

3.6.1.2 Review potential SP and RP variables

Once the sociodemographic and household characteristics of the sample have been explored, it is time to assess the sample's consumer and travel behavior and preferences by performing a descriptive statistical analysis of both the SP and RP variables. The modeler should develop

descriptive statistics of any and all variables of heightened interest, that is, those that directly relate to the *topic of interest* that the survey was designed for (as decided in Section 3.1.1).

For example, for the WPL survey, the topic of interest is how workplace location choices and preferences changed across four different time periods: pre-COVID, during the peak of COVID prior to vaccination, currently (as vaccines have become widely available), and in a post-COVID future. Specifically, this study investigated where employees may prefer to work (given three different options) in the unpredictable future. When looking at the resulting data, the analyst first developed descriptive statistics (using single-variate frequency tables and multivariate crosstabs with R database management software) of every variable in the dataset that related to workplace location. This included the following variables, which have their own unique value during each of the four time periods:

- Commute characteristics
- All workplace location choice, splits, and trends
- All workplace environment characteristics

Table 3 is an example of a univariate table from this survey. Table 3 displays three columns of RP data ("Before COVID," "During COVID," and "Now") and one SP data column ("In the future").

Table 3: Univariate Example: Time Period and Frequency of Teleworking

How often did/do/will you telework	Before COVID		During COVID		Now		In the future	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Never telecommuted	582	59.4%	84	8.6%	334	34.1%	359	36.6%
A few times per month	161	16.4%	82	8.4%	133	13.6%	139	14.2%
Once per week	58	5.9%	36	3.7%	71	7.2%	70	7.1%
2–4 days per week	68	6.9%	123	12.6%	195	19.9%	234	23.9%
5 days a week (every day)	111	11.3%	655	66.8%	247	25.2%	178	18.2%

While Table 3 shows a lot of important trends, it is important not to overanalyze them or analyze an excessive many RP and SP potential outcome variables during the descriptive statistical analysis step, as it may an overwhelming effort and loss of time. Pick a few important variables to explore and then a few important aspects of each developed table that best embody and reflect trends about the topic of interest and, if applicable, the variables that will be the main outcome variables during the choice modeling process.

For example, some important trends to notice in Table 3 are condensed as follows:

- Roughly 50% of the overall sample shifted away from never teleworking and 50% shifted to teleworking every day in the during-COVID (before vaccination) time period compared to before COVID.
- Relative to before COVID, current rates of teleworking once or more per week (in the "Now" column) are much higher across all frequencies, whereas never or rarely teleworking has decreased.
- The current and forecasted future telework trends revealed by the table are roughly identical. The main difference lies between the "2–4 days per week" and the "5 days a week (every day)" telework frequencies; in the SP choice occasion, more respondents selected the former group than the latter, relative to today's RP data.

This descriptive analysis helps determine the variables to be used in choice modeling analysis. However, in the example of the WPL study, the main outcome variable was not any of the variables presented in Table 3. While there is a column of SP data in Table 3, more data from the SP experiment will be used for the choice modeling example in Section 3.6.2.

Additional descriptive statistics, beyond frequency and crosstabs, may be necessary for data directly related to an SP experiment, especially if the experiment was not a discrete choice. Additional descriptive statistics may include:

- The mean value of a continuous alternative
- The range of values
- The minimum or maximum

For the WPL study, descriptive statistics, which can be found in Table 4, were calculated for the portion of choice occasions with positive participation[4] for each WPL alternative.

Table 4: Example of an Additional Descriptive Statistic: Portion of Choice Occasions with Positive Participation

WPL Location	Total number (%) of choice occasions with positive participation	
	RP Data (1,136 total)	**SP Data** (1,136*2 total)
Home	671 (59.1)	1635 (72.0)
Work Office	897 (79.0)	1561 (68.7)
Third WPL	86 (7.6)	330 (14.5)

The statistics displayed in the RP data column of Table 4 indicate where the respondents worked in the month prior to responding to the survey. In the past month, working in the office at least

[4] In the WPL SP experiment, respondents were asked to allocate a month of workdays across all three WPLs. For example, an employee who reports to work 22 days in a moth, would have 22 days to either split up across all three WPLs, across only two WPLs or assign all their time to a single WPL. Positive participation implies that the respondent assigned at least one day to that WPL.

once was about 20% more common than working from home at least once, while working from a third workplace was significantly less common, with only 7.6% of respondents doing so. The SP data column of Table 4 is the positive participation in each workplace alternative in response to the scenarios presented in the SP experiment. Here, the home and work office alternatives are chosen at least once at similar rates, with around 70% participation, while the third workplace location is less likely to be chosen for at least one day a month, at only 14.5%.

3.6.1.3 Ensure the SP variables are grounded with RP data

Before choice modeling can occur, the analyst must determine if they will model RP-SP data jointly or just the SP data. Neither is inherently better, but depending on which is selected, the modeler may have to take an additional descriptive statistical analysis step. The goal of this step is to ground the SP variables with RP data in order to acknowledge and limit bias in the SP data.

If joint RP-SP modeling is selected, then the modeler can proceed directly to the next step, as joining the RP variables with the SP variables will acknowledge and limit the bias of the SP variable. The main limitations of SP data:

- "Setting bias" (the choice is made in a hypothetical setting)
- "Policy bias" (respondents attempt to influence the outcome)

Therefore, ensuring that responses to the SP experiment are similar to those in comparable RP questions will ensure limited bias of the SP data.

If only SP modeling is selected, then the modeler must pay close attention to the descriptive statistical analysis performed in Section 3.6.1.2 (particularly the analysis where RP and SP data is directly compared, as is done in Table 4) to ensure that the SP variables follow similar trends to the RP variables. This comparison is necessary to make sure that the SP data, which is reflective of an individual's response or behavior in a hypothetical and hard-to-imagine scenario, is realistic. This realism can be confirmed by descriptively comparing the SP data with a similar variable measured in a current context through RP data.

For the WPL study, only a SP variable was analyzed through a choice model, therefore a descriptive confirmation of the limited bias in the SP data is required. This was done through comparing the RP data from respondents' current monthly workday split across the three workplace alternatives with the SP data on monthly workday split in response to the scenarios presented in the SP experiment, as presented in Table 4. We can see that the RP and SP data are not identical, but we can confirm limited bias by assessing that the general trends across and between all alternatives are similar, and no alternative has a significantly different representation in the SP data relative to the RP data (an indication of greater bias would be, for example, if the third workplace in the SP data had over 50% participation, compared to 7.9% in the current RP data).

Additionally, to make the comparison and ground the SP data, the analyst must consider the differences in the situational circumstances between the SP and RP data. In the WPL study, the RP questions in the survey were not controlled for distraction level or COVID risk, while the point

of the SP experiment is to study how variations in workplace environment and location characteristics impact employee workplace location preferences. Therefore, the slight difference in the RP and SP data for positive participation for working from home is excusable, and the SP data can be deemed limited in bias and grounded through adequate comparison with RP data. Similar considerations, assumptions, and assessments should be evaluated whenever RP data is directly compared to SP data.

3.6.2 Choice modeling analysis

The next step of the analysis process is to develop a choice model of the RP-SP, or just SP, outcome variables. Choice modelling aims to reflect the decision process of an individual or segment of the population via revealed or stated preferences made in a particular context or contexts. Choice models use discrete, continuous, ranked, ordered, or other formats of choices or preferences in order to deduce the sample and population's positions on the topic of interest on a relevant latent scale.

3.6.2.1 Design an analytic framework

Before beginning the choice modelling process, it may be helpful to design an analytic framework to determine how to best answer the question being posed in the SP experiment, as well as to establish how to effectively incorporate the sociodemographic data and the answers to other related RP questions. An analytic framework's goal is to create a visualization of the exogenous variables (sociodemographic variables, built environment variables, or other individual or household attributes reported through RP questions) and their link to the main outcome variables, including the SP experiment and the RP questions, which will be jointly modeled. An example of an analytic framework is shown in Figure 3. In this analytic framework, the RP data is bordered by dotted lines, while the SP data boxes are contained in solid lines.

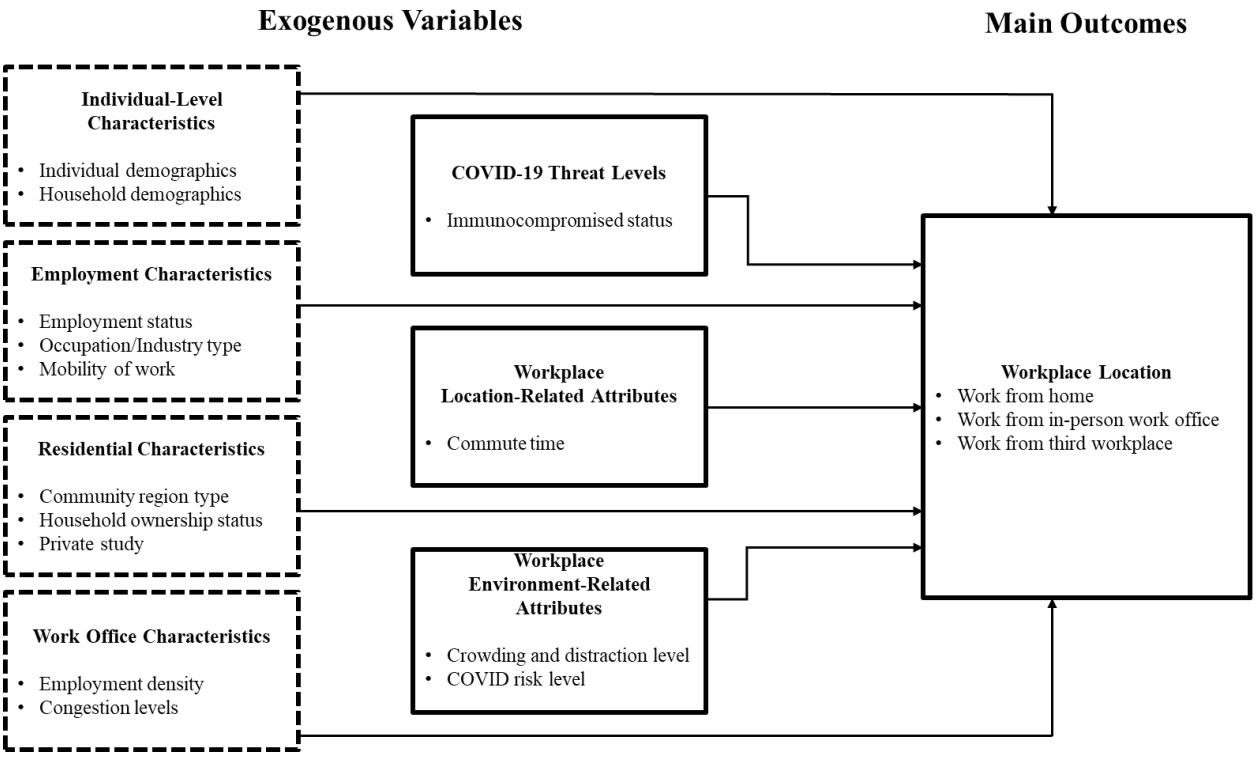

Figure 3: Analytic Framework

3.6.2.2 Model the data

After decisions have been made about what model to use and how it will be set up, the bulk of the modeling and analysis of the data begins. Most of the analytical tools that may be used for forecasting travel behavior based on RP survey data can also be used with SP survey data. These analytical tools include frequency tabulations, linear regressions, discrete choice models, and ordinal variable models. However, some analytical tools are generally better suited for use with SP components. Examples of these analytical models are ranking models and best-worst preference models, which can be used only if the preference elicitation method allows. By jointly modeling the SP and RP components or through the descriptive comparison of hypothetically based SP data with realistic and true RP data, the modeler can "prove" realistic representation and "disprove" bias in the SP data so that it can be confidently used on its own when modeling. Some effective methods for controlling bias in this manner are demonstrated in Bhat and Castelar (2002). When designing a survey, researchers ask certain RP questions before and after the SP portion of the experiment so they can serve as anchors to ensure that the SP responses (made in the context of hypothetical scenarios) are reasonably consistent with the actual travel behaviors manifested by individuals, revealed in RP responses (Loomis, 2011), which aids in "proving" realistic representation and "disproving" bias of the SP data. RP questions should be curated with this specific analysis intention in mind; their use alongside the SP data ensures "grounding" of the SP data with the RP component.

Although each of the RP grounding questions may provide valuable data in its own right, anchoring the SP components with RP data results in a more robust analysis of travel behavior and allows projection into a future environment quite different from today's reality. The objective of these RP questions is simple: to provide data for the continued development and refinement of travel demand models. Recognizing the actual travel habits and preferences of respondents is vital. Therefore, in an RP-SP choice model, the RP data is estimated as a dependent outcome that is jointly modeled alongside the SP experiment's dependent outcomes, while also informing the SP experiment's dependent outcomes as an endogenous explanatory variable. In an SP-only choice model, the RP data is employed as an array of exogenous variables while estimating SP dependent outcomes.

Applications of joint RP-SP models and SP-only have become increasingly popular in transportation research. One of the most common modelling structures is a *mixed multinomial logit formulation.* A mixed multinomial logit model provides a straightforward method to consider both the RP and SP responses for a single individual simultaneously. This formulation relies on two components: *mixed* and *multinomial.*

> **Mixed:** The presence of multiple observations of stated choice responses and actual revealed behavior for each sampled individual suggests that the potential for correlated responses across observations is a violation of the "independence of observations" assumption in classical model estimation. The *mixed* formulation relaxes the independence assumption and accounts for the correlation in decision-making across multiple choice instances for the same individual. In a *mixed* logit model, all parameters or variables are assumed to vary from one individual to another, therefore accounting for the heterogeneity of the population.
>
> **Multinomial:** In any single SP or RP question, an individual can be presented any number of options to choose from. A *multinomial* formulation is used when the decision in question is nominal (or categorical, meaning that it falls into one of a set of categories), rather than binary (offering only two possible responses). For example, in the workplace location choice occasion, the respondent is presented three options: work from home, work at the office, or work from a third workplace). Including these three options creates a multinomial regression setup.

Essentially, a mixed multinomial logit model takes the structure of a multinomial logit model for each individual, conditional on the coefficient value (taste sensitivity to variables) for that specific individual. This coefficient value may be affected by unobserved individual-specific factors; for example, some people, because of their sociable and extroverted nature (which would be an unobserved variable in most studies), may intrinsically prefer to go their workplace rather than work from home. The effect of such unobserved individual factors (in terms of shifting taste sensitivity) is assumed to be captured in a realization from a specific mixing distribution (typically a normal distribution). Finally, the analyst simply integrates the mixing distribution (with the multinomial logit kernel as the conditional basis) to get the desired probability in the mixed multinomial logit model. For further context, the modeling type used to estimate the framework

presented in Figure 3 was a version of a multinomial model known as the multiple discrete-continuous extreme value (MDCEV) model (Bhat, 2008). A MDCEV model is a beneficial approach to use when estimating individuals' behavior, as it includes the ability to measure both a preference (the discrete variable) and the frequency of choosing that preference (the continuous variable) in order to provide a multidimensional analysis of the issue in question. This is especially helpful when projecting behavior in an unprecedented future or reaction to an emerging technology, as more information and more insight can be extracted from the SP data at hand.

Numerous statistical software programs can estimate mixed multinomial logit models; many of these, such as R, are free or open-source. Other programs, such as SPSS, Stata, and Gauss, can also be used to estimate a mixed multinomial logit model, though they require a yearly subscription to access.

3.6.2.3 Interpret the choice model results

As the data is modelled, the results must be appropriately interpreted, as is the case throughout the analysis process. Interpreting the results from a jointly modelled RP-SP formulation is identical to any other regression analysis of a similar model. As discussed in the previous step, a common modeling strategy employs a mixed multinomial logit model. Compared to modeling the RP and SP data individually, there is no difference in terms of the interpretation of the results, though joint modeling beneficially impacts the estimate/coefficient value by influencing correlation and error effects. However, there are immense and comparative benefits, for methodological and application purposes, to modeling the SP data by itself.

Table 5 and Table 6 present two examples of table formats that may be used to clearly display the results from the model to make interpretation more straightforward. Notice in both tables each variable has a "coefficient" value and a "t-stat" value. The numeric value of the coefficient is not important, but rather the sign (whether it is positive or negative). If the coefficient is positive, it means that, relative to the base outcome and the base for that variable, respondents in that variable category are more likely to choose the option in question, for example a specific workplace location. If the sign is negative, respondents are *less* likely to choose that specific workplace location. Certain exogenous variables should be kept in the model only if their t-stat value is above either 1.5 or 2, depending on the analyst's choice of significance level. If the value of the t-stat (regardless of sign) is lower than the designated significance level for a certain outcome variable, then that variable is omitted from that specific outcome, replaced in the table with a long dash (—). This then implies that the variable has no significant impact on that outcome, relative to the base outcome for the same dimension. As stated earlier, the process used to arrive at the interpretation of results displayed in Table 5 and Table 6 was identical to that used by a logit model employing only RP data (whether mixed, not mixed, multinomial, or binary). Table 5 presents the skeleton of a table for estimating results for a joint RP-SP analysis, while Table 6 presents an example of an SP-only estimation (which will be the structure used in other associated project documents, such as Technical Memorandums 8, 9, 10, and 11, as well as in the Final Report). Some preliminary estimation results will be explained following Table 6 as an example of analysis of an SP-data-only model (whether or not it's accompanied by an RP main outcome, as is the case for Table 6).

Table 5: Example of Joint RP-SP Table

Exogenous Variables (base category)	Pre-COVID Workplace Location Choice (RP)						Post-COVID Workplace Location Choice (SP)					
	Work from Office (Base)		Telework		Hybrid of Both		Work from Office (Base)		Telework		Hybrid of Both	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Individual-Level Characteristics												
Gender	--						--					
Age	--						--					
Education Level	--						--					
Employment Status	--						--					
Household Characteristics												
Income Level	--						--					
Presence of Children	--						--					
Household Structure	--						--					
Vehicle Availability	--						--					
Residential Location	--						--					
Commute and Workplace Characteristics (pre-COVID and during COVID)												
Commute Time												
Commute Distance												
Teleworking Status	--						--					
Workplace, Home, and Job Attributes												
COVID Risk Level	NA		NA		NA		--					
Distraction Level at Home	NA		NA		NA		--					
Change in Commute Time	NA		NA		NA							
Splitting/Shifting Work Hours	NA		NA		NA							
Level of Crowding at the Out-of-Home Workplace	NA		NA		NA							
Workplace Safety Implementation for COVID	NA		NA		NA							
Pre-COVID Workplace Location Choices (RP)												
Telework	NA		NA		NA		--					
Hybrid of Both	NA		NA		NA		--					

Table 6: Example of SP Table with Estimation Results

Exogenous Variables (base category)	Workplace Location Choice					
	Work from Home		Work from the Work Office		Work from a Third WPL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Individual-Level Characteristics						
Gender (male)						
Female	--		-0.200	-1.50	--	
Age (18 to 29 years old)						
30 to 64 years old	--		0.550	2.50	--	
65 and older	--		0.600	2.20	--	
Household-Level Characteristics						
Household Income (<$100,000)						
$100,000 to $249,999	--		-0.410	-4.80	--	
≥$250,000	--		-0.460	-3.20	-0.450	-1.90
Geographic WPL Attributes						
Commute Time						
Commute time	NA		-0.750	-2.70	-0.750	-2.70
Population Density of the Residence (rural)						
Suburban	--		-0.700	-6.00	-0.500	-2.80
Urban	--		-0.900	-7.50	-0.500	-2.80
Environment WPL Attributes						
Distraction Level (low)						
Medium	-0.140	-1.70	-0.600	-3.00	-0.580	-3.80
High	-0.700	-2.30	-0.600	-3.00	-0.750	-4.00
Baseline Preference Constant	NA		1.330	4.60	-0.390	-1.90

The results provide in Table 6 are a simplified version of the real, estimated model (which can be found in the working paper Asmussen et al. (2022), and in the final report associated with this project). The main outcomes for this model are derived directly from the WPL SP experiment, which was developed alongside this guidebook. In this example, we have modeled employees' preference for working from each of the WPL alternatives (the discrete dimension of the MDCEV model) and the frequency of working from each of the WPL alternatives (the continuous dimension of the MDCEV model). For simplicity's sake, we have excluded review of the frequency dimension, or the satiation parameter values, that are associated with the estimation of a MDCEV model.

Estimation interpretation tips

For the sake of efficiency and concision in this guidebook, we will not review all of the results from Table 6. A few results have been selected for explanation in order to emphasize the key analysis approaches, which are referred to as *Interpretation Types*. And in these instances, only the value and sign of certain coefficients will be discussed, rather than their implications. Additionally, note that a "tagged alternative" refers to the alternative for which the variable is estimated. The six main interpretation types are as follows:

- Interpretation Type 1: Single estimated variable, single alternative

    - This is the simplest type to interpret. Only the sign of the estimated effect is important for analysis.
    - Example: Gender variable
        - Base: Male
        - Estimated variable 1: Female for tagged alternative *Work from Work Office*
            - Estimated effect: Negative
            - Interpretation: women employees, relative to males, have a lower preference for working from a Work Office and a higher preference for either remote option.
- Interpretation Type 2: Multiple estimated variables, single alternative
    - Here, it is important to analyze both the sign and the value of both estimated effects.
    - Example: Age variable
        - Base: 18 to 29 years old
        - Estimated variable 1: 30 to 64 years old for tagged alternative *Work from Work Office*
            - Estimated effect: Positive
            - Interpretation: Employees aged 30 to 64, relative to younger employees, have a higher preference for working from a Work Office and a lower preference for either remote option. However, their preference for the in-person office is lower than that of estimated variable 2 immediately below, 65 years and older (coefficient of 0.550 compared to 0.600).
        - Estimated variable 2: 65 years and older for tagged alternative *Work from Work Office*
            - Estimated effect: Positive
            - Interpretation: Employees aged 65 and older, relative to the youngest employees, have a higher preference for working from a Work Office and a lower preference for either remote option. Their preference for the in-person office is the highest of any age group, and their preference for a remote WPL is the lowest for any age group.
- Interpretation Type 3: Multiple or single estimated variables, multiple alternatives
    - The difference between this interpretation type and the previous two is the inclusion of multiple alternatives, instead of just one. In this case, analysts can only compare the tagged alternative to the untagged alternatives, or in the case of the example models, the only untagged alternative. They cannot compare the value across alternatives when there is more than one tagged alternative, only the sign.
    - Example: Income variable
        - Base: Less than $100,000
            - Estimated variable 1: $100,000 to $250,000 for tagged alternative *Work from Work Office*
                - Estimated effect: Negative

        - Interpretation: Relative to those households making under $100,000, employees in households making between $100,000 and $250,000 a year have a lower preference for working from an Work Office and a higher preference for either remote option (this is the same interpretation format as Type 1, because there is no second alternative for this middle-income level).
    - Estimated variable 2: $250,000 or higher for tagged alternative *Work from Work Office*
        - Estimated effect: Negative
        - Interpretation: Relative to those households in lower income groups, employees in households making between $100,000 and $250,000 a year have a lower preference for working from an Work Office and a higher preference for working from home. Note that the only alternative for comparison here is "work from home," which has no tag in Table 6. This is due to the inclusion of a second tagged alternative for this estimate variable.
    - Estimated variable 3: $250,000 or higher for tagged alternative *Work from Third Workplace*
        - Estimated effect: Negative
        - Interpretation: Relative to those households in lower income groups, employees in households making more than $250,000 a year have a lower preference for both working from an Work Office and a third workplace.
- Interpretation Type 4: Multiple alternatives, but the same estimate value and sign for all (generic variable)
    - This is the case of generic variables; usually this means analysts are trying to measure the marginal utility of a variable such as time or money. This can be done for both discrete, continuous, or dummy variables.
    - Example: Commute time
        - Base: no base, continuous variable
        - Estimated variable 1 and 2: Commute time for both tagged alternatives, *Work from Work Office* and *Work from Third Workplace*
            - Estimated effect: Negative
            - Interpretation: As the commute time rises for either outside-of-home alternative (there is no commute time when working from home, the third alternative), the preference for that alternative decreases.
- Interpretation Type 5: Multiple estimated variables, but the same estimate value and sign for all estimated variables for a single alternative, but there are multiple alternatives

    - Here, the estimate values are the same because variable levels for one alternative have been condensed as they have separate effects for the other alternatives. This interpretation is performed for its simple visualization benefits.
    - Example: Population density of the residence
        - Base: Rural
        - Estimated variables 1 and 2: Suburban and Urban for tagged alternative *Work from Third Workplace*
            - Estimated effect: Negative for both
            - Interpretation: Relative to employees who live in a rural area, employees in a suburban or urban area will have the same preference against working from a third workplace and a higher preference towards working from home.
        - Estimated variable 3: Suburban for tagged alternative *Work from Work Office*
            - Estimated effect: Negative
            - Interpretation: Refer to Interpretation Types 2 and 3.
        - Estimated variable 4: Urban for tagged alternative *Work from Work Office*
            - Estimated effect: Negative
            - Interpretation: Refer to Interpretation Types 2 and 3.
- Interpretation Type 6: Multiple or single estimated variables, across all alternatives
    - In order for there not to be an untagged alternative (or a base alternative to compare the trends in the other alternative), the estimated variables must be different across all alternatives, or must be alternative-specific variables. Being female, for example, is not an alternative-specific variable, but distraction level, which varies across alternatives, is. In this case, each alternative can only be compared to the variable base (such as comparing "high distraction level" to "no distractions"), instead of between or across alternatives (as is done when comparing the impact being a female has on choosing one WPL over another).
    - Example: Distraction Level
        - Base: Low distraction level
        - Estimated variable 1: Medium distraction level for tagged alternative *Work from Home*
            - Estimated effect: Negative
            - Interpretation: Relative to a low distraction level, when the distraction at the home office is at a medium level, an employee will be less likely to work from home. However, they will be more likely to work from home when the distraction level is medium as compared to high (for the same reasoning as used in Interpterion Type 2).
        - Estimated variable 2: High distraction level for tagged alternative *Work from Home*

    - Estimated effect: Negative
    - Interpretation: Relative to the two lower distraction levels, when the distraction at the home office is at a high level, an employee will be less likely to work there.
  - Estimated variables 3 and 4: Medium and high distraction levels for tagged alternative *Work from Work Office*
    - Estimated effect: Negative
    - Interpretation: Use Interpretation Type 5.
  - Estimated variables 5 and 6: Medium and high distraction levels for tagged alternative *Work from Third Workplace*
    - Estimated effect: Negative
    - Interpretation: Same as for estimated variables 1 and 2 for this same Interpretation Type.

These six main interpretation types for SP and RP-SP estimates are not unlike those that are used for RP-only purposes. The only difference is that analysts are able to predict future behavior (through an SP experiment), as opposed to modeling what respondents already do.

3.7 Integration of Results and Data

This section describes how the results may later be integrated into the four-step travel demand model that is used in traditional travel demand analysis methods:

- *Trip Generation*: Through the use of SP questions, the frequency of origins and destinations of trips in a particular zone and for a particular purpose can be predicted. Matching respondents' answers with their individual and household demographics, as well as their socioeconomic situation, will provide a type of regression tool for the prediction of future trip generation.
- *Trip Distribution*: The motivation behind each trip will match origins with destinations—often through use of a gravity model.
- *Mode Choice*: The proportion of trips between origins and destinations via specific modes calculated using a logistic (logit) model will predict the probability of a particular respondent's future mode choice.
- *Route Assignment*: The allocation of trips between origins and destinations by a particular mode and route is predicted based on the shortest travel time path. However, this step is typically challenging to perform, as travel time is dependent on demand, yet demand is dependent on travel time (bi-level problem). SP surveys provide a tool to predict demand, ending this loop and anticipating route assignment scenarios.

More generally, SP questions, once designed and thought through, can be added rather easily toward the end of any regional travel survey instrument. Designed appropriately, and carefully, it would only add about two or three additional questions. It is true, however, that such questions

would take longer to respond to than a quick RP-based question, because respondents have to go through the scenario description, understand what the attributes mean, and also absorb the level of each attribute being presented. And then there is the choice to be made, as opposed to the choice that already was made in an RP context.

The "trick" in SP question presentation is to present the scenario to respondents in clear and plain-speak that is not too technical. Sometimes, researchers, who are aware of many nuances of a possible future travel environment, may want to get too specific or may worry about presenting things with ambiguity if all the minutest details of the scenario are not made clear. They may also want to include a whole bunch of attributes to be as specific as possible about the experimental context, and may desire a fine categorization of the attribute levels. However, individuals in society will not all be transportation experts, so travel demand practitioners and travel survey designers have to simplify and minimize material so as not to overwhelm respondents with too much information. It is okay to leave some room for respondents to interpret some aspects of the scenario without getting into too many specifics. This obviously calls for a balancing act between leaving too much ambiguity and being too specific. Overall, the scenario description (and the descriptions of the accompanying attributes and attribute levels) should be concise, non-technical, and written in a way that is easy to absorb for respondents. In this regard, it is critical to conduct a "friends-and-family" pilot survey, and potentially multiple rounds of these, to get the scenario description just right. Besides, such pilot surveys may also identify attributes that may be relevant, but were not considered in the original design because transportation practitioners and researchers may have missed out on some broader aspects of daily life that affect the context of the study.

4. Conclusion

As part of the ongoing RP-SP project, TxDOT embarked on an effort to consider the inclusion of SP questions and an SP experiment within their usual RP-based survey procedures. While RP data reflects respondents' actual current or past behavior, SP data is data collected about choices that respondents *would* make in different scenarios that are drawn up for the respondent to picture themselves in. TxDOT decided to develop guidelines for SP survey data collection on current and future activity travel patterns. This guidebook discusses the design of an SP survey and how the results may be integrated into the traditional analysis method.

The first step is to determine the topic of interest that will lead to insightful conclusions that are relevant to policy. Many topics of interest exist, and more will emerge upon future findings. It must be noted that there is a certain amount of risk and uncertainty attached to different SP experiment topics. Acknowledging this uncertainty is imperative, as SP results can indicate a likely trend but cannot guarantee that future behavior will follow their predictions. Various strategies can be used to reduce this uncertainty.

Next, the details of survey format and survey elicitation mechanisms should be established. The word choice for the descriptions of the alternatives, attributes, their levels, and the entire SP experiment must be exact and deliberate, with a focus on being clear and concise. SP questions are then framed with RP questions; the survey deployer chooses those that will be most beneficial and

logical to incorporate. After linking the RP and SP portions of the survey, a pilot should be deployed to ensure the flow and readability of the survey questions and format. Once the pilot has led to a satisfactorily revised survey, the survey can be distributed to its full target audience. Finally, once an appropriate sample size is achieved, the data is organized, analyzed, and modelled, and the results are integrated into current analysis.

SP surveys can be valuable when designed, deployed, and applied properly. The resulting data can be used to predict future conditions and behaviors, providing insights that can be applied to timely policy updates. The results of an effectively designed and deployed RP-SP survey may point to a new direction for potential investigation. This guidebook addresses RP-SP survey development in a step-by-step fashion, from developing the scenario and reducing risk through the integration of the resulting data. From this guidebook, an RP-SP survey may be developed on any topic of interest.

5. References

Asmussen, K. E., Mondal, A., Bhat, C. R., and R.M. Pendyala (2022). On Modeling Future Workplace Location Decisions: An Analysis of Texas Employees. *Technical Paper,* Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, *42*(3), 274–303.

Bhat, C. R., and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B*, 36(7), 593–616.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, 25(2), 363–370.