Article title: Preprint: Mobility surveys beyond stated preference: Introducing MyTrips, an SP-off-RP survey tool, and results of two case studies

Authors: Christian Rudloff[1], Markus Straub[2]

Affiliations: AIT Austrian Institute of Technology GmbH[1]

Orcid ids: 0000-0001-5424-8560[1]

Contact e-mail: christian.rudloff@ait.ac.at

License information: This work has been published open access under Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Conditions, terms of use and publishing policy can be found at <https://www.scienceopen.com/>.

Preprint statement: This article is a preprint and has not been peer-reviewed, under consideration and submitted to ScienceOpen Preprints for open peer review.

Funder: This work is part of the scientific project "SynArea II" supported by the Austrian Research Promotion Agency.

DOI: 10.14293/S2199-1006.1.SOR-.PP0T28V.v1

Preprint first posted online: 28 June 2021

Keywords: Hypothetical bias, Intermodal transportation, Micro transit, Mode Choice modeling, SP-off-RP, Transport data collection

RESEARCH

Preprint: Mobility surveys beyond stated preference: Introducing MyTrips, an SP-off-RP survey tool, and results of two case studies

Christian Rudloff* and Markus Straub

*Correspondence:
christian.rudloff@ait.ac.at
AIT Austrian Institute of
Technology GmbH, Giefinggasse 2,
Vienna, Austria
Full list of author information is
available at the end of the article

Abstract

When introducing new mobility offers or measures to influence traffic, stated preference (SP) surveys are often used to assess their impact. In SP surveys respondents do not answer questions about their actual behavior but about hypothetical settings. Therefore, answers are often biased. To minimise this hypothetical bias, so-called stated preference-off-revealed preference (SP-off-RP) surveys were developed. They base SP questions on respondents' revealed behavior and place unknown scenarios in a familiar context. Until now this method was applied mostly to scenarios investigating the willingness to pay. The application to more complex mode or route choice problems, which require the calculation of routes, has not yet been done.

In this paper, the MyTrips survey tool for the collection of SP-off-RP data based on respondents' actual mobility behavior is presented. SP questions are based on alternatives to typical routes of respondents, which are calculated on the fly with an intermodal router. MyTrips includes a larger survey and collects mobility diaries for one day representing respondents' daily routine, calculates alternative routes and creates SP questions based on a Bayesian optimal design.

Results from two case studies investigating behavior changes are presented. The first case study investigated the extension of a subway line in Vienna, Austria. The second case study focused on the introduction of micro transit vehicles in a rural setting, replacing infrequent bus services. Results of the two case studies show a difference in response behaviour between SP and RP settings and suggest a reduction of hypothetical bias. For the latter study a Latent Class SP-off-RP model was estimated. It shows that availability and accessibility of public transport are the main influence on the willingness to use it independent of other household characteristics.

Keywords: Hypothetical bias; Intermodal transportation; Micro transit; Mode choice modeling; SP-off-RP; Transport data collection

1 Introduction

Introducing new mobility options is costly and acceptance of new measures by travellers is hard to estimate in advance. Surveys are the prime source of data for impact assessment. In particular, stated choice experiments are used to collect data about hypothetical scenarios and choice models are applied to forecast adoption of new travel choices.

In general, there are two kinds of data sources for the estimation of mode and route choice models. The first are revealed preference (RP) data where actual behavior of travelers is recorded. The second are stated preference (SP) data, where

respondents are asked to choose between different scenarios in a survey setting. Both have advantages and disadvantages. The main advantages of SP data are that hypothetical scenarios can be included and that one has control over the survey design. RP data on the other hand is real behavior and does not suffer from hypothetical biases which is often a problem in SP data where people must answer questions about hypothetical behavior in an unknown setting. The problem of hypothetical biases is well studied, in particular in the area of willingness to pay. Several papers show that hypothetical biases exist, e. g. [1, 2, 3] show that there is a discrepancy between stated choice experiments and revealed preferences. In [4] a similar comparison between a SP experiment and GPS data is shown but the paper also shows that hypothetical bias is reduced by mitigation techniques like showing a cheap talk script to respondents that informs them about the problems that arise from hypothetical bias before they answer the SP questions.

Other techniques combine different data sources, mainly SP and RP data. Pivoting techniques minimise the hypothetical bias by basing the design of the SP alternatives on a recent route chosen by the respondents in real life (see e. g. Rose et al. [5]). Pivoting techniques were applied in a series of papers [6, 7, 8] by Hensher and Rose where methods for designing stated choice (SC) experiments from pivot alternatives collected in a computer assisted personal survey were developed. The pivot alternatives were subsequently adjusted according to a D-efficient design but no actual routing step was applied in the survey design. In [9] the number of transfers in the SP-questions to determine transfer penalties of passengers were based on the real number of transfers in a collected RP-route.

In a series of papers Train and Wilson [10, 11] develop the Stated Preference-off-Revealed Preference (SP-off-RP) approach, where an actual revealed choice situation is offered within the SC experiment by changing one or more characteristics but leaving the main part of the choices as in the real situation of the respondent. In addition, they introduce a choice modelling framework to handle the resulting SP-off-RP data.

In this paper, the novel online survey tool MyTrips is presented. MyTrips allows the collection of SP-off-RP data based on routing. This SP-off-RP data offers the possibility to reduce hypothetical distortion for transport and route choice problems. In addition we present results of two case studies where MyTrips was applied. The main contributions of this paper are:

- C1** Methodologies applied in the MyTrips survey tool for an SP-off-RP survey based on routing:
 - (a) Novel online survey collection methodology
 - (b) Calculation of route alternatives based on collected trip diaries
 - (c) Design of SP choice sets from route alternatives
- C2** SP-off-RP evaluation methodology using Latent Class Models
- C3** Two case studies applying the MyTrips survey tool and corresponding evaluation methodologies

While the SP-off-RP data collection methodology was applied in a number of cases, these cases had in common that the RP setting was relatively easy, making the creation of the SP alternatives possible. In the original paper Train and Wilson [11] use the example of shippers and alternative routes for their goods. The SP

alternatives are pivoted on the chosen alternative by changing the tariff or travel time. Other examples within the transportation sector are Arellana et al. [12] where departure time choice of workers in Santiago was collected and SP questions were designed with different attribute levels of departure choices and changes in mode choice behavior. In Cranenburgh et al. [13] vacation choices and possible alternatives are collected in the RP stage and in the SP stage random pairs with changed costs and travel times were offered as SP alternatives. Yu et al. [14] study rebound effects caused by energy efficiency in cars by applying an SP-off-RP survey where they ask about the current vehicle in a household and replace the annual operating costs automatically for a more efficient replacement.

For none of the cases, where SP-off-RP settings were used for a survey, a routing step was necessary for the setup of the SP survey. However, when changing larger parts of the transportation system a good SP-off-RP setting is harder to achieve since the questions need to be put into a familiar setting and a routing step is added to achieve a familiar setting for the SP-route. This connection of router and online surveying tool is technically more challenging. Furthermore, the survey design is getting more difficult due to the added routing step preventing direct manipulation of variable levels in the SP design. Some design strategies are already present in literature for pivoted SP questions. In Rose et al. [5] a D-efficient design is tested where the experimental design was constructed based on assumed population level averages for each of the design attributes. Arellana et al. [12] suggests a five step method based on the Bayesian efficiency criterion where a generic SP design is found first that is customised to the revealed choices for each respondent.

In addition to the collection tool, this paper also presents methodologies to analyse the collected SP-off-RP data and shows results for two case studies. Besides general survey analyses, latent class models (LCM) are applied for a more detailed examination of respondents' mode choice behaviour. For a good introduction of LCM see e.g. [15]. For an incorporation of SP-off-RP data into choice model estimation see e.g. [11]. While other approaches for personalising models exist, LCM might be more suitable when personalising choice models for single respondents (see e.g. [16] or [17]) since other methods need data for each respondent for personalisation. The model presented in this paper is meant to serve as a mode choice model in an agent based simulation model. Therefore, only socio-demographic information is available in the synthetic population. The LCM approach can be used to apply this data for personalising the routing model for each agent.

The remainder of this paper is organized as follows. In Section 2 we present the MyTrips Methodology for data collection and data analysis. Section 3 describes two case studies conducted with MyTrips. Section 4 presents the results of the two case studies. Finally, we present our conclusions and necessary future work.

2 MyTrips Methodology

The goal of MyTrips is to improve online (web-based) mobility surveys. The main cases of application are surveys for assessing the impact of (future) mobility scenarios on mode and route choice, i.e. which modes of transport and routes people prefer to use for their everyday mobility.

In the following sections we describe the MyTrips concept, the main methodological and technical contributions of the MyTrips survey tool, as well as first approaches for analysing collected SP-off-RP data.

2.1 The MyTrips Data Collection Methodology

The main novelty of MyTrips is real-time personalisation of questions, i. e. routes, in the SP-off-RP survey to improve quality and accuracy of the hypothetical SP questionnaire: For RP respondents' preferences are inferred through the calculation of alternative routes that would have been available to them at the time. For SP the personalisation of routes helps respondents to better imagine the survey scenario as alternatives' setting is familiar. A MyTrips survey consists of the following parts, as shown in Figure 1:

- P1** Introductory questionnaire
- P2** Trip diary
- P3** Choice set calculation (consisting of alternative routes)
- P4** Main questionnaire
- P5** Personalised SP-off-RP questions (one for each choice set)

The parts **P1** and **P4** together form the underlying conventional online mobility survey. In **P1** first introductory questions are posed to filter respondents not meeting the survey requirements, e. g. living outside of the study region, and to collect preferences required for the calculation of alternative routes such as the ability to cycle or availability of personal vehicles. **P4** represents the major part of the conventional survey consisting e. g. of questions about mobility preferences or demographic data as well as an introduction to the hypothetical scenario and first pure SP questions about expected behaviour.

MyTrips builds upon these parts and adds **P2** as RP data and **P3** for corresponding SP-off-RP questions and finally **P5** for a stated choice experiment.

In Figure 2 the user interface for the collection of trip diaries in **P2** is shown.

For each trip respondents must fill in start and end location, departure and arrival times, and mode of transport. Intermodal trips are split into single-mode stages. Depending on the survey goal and scenario, the degree of detail required for the modes of transport may vary. For one survey it may be sufficient to provide a single trip for a complex public transport journey with several line changes. For another survey line and changeover information or a differentiation between car driver and passenger may be required. In addition, the activity between trips, e. g. being at work, shopping, or spending leisure time, is inquired.

Choice sets for **P5** are calculated based on **P2**, where each choice set consists of two or more routes with the same start and end location. The detailed calculation methodology is presented in subsections 2.2 and 2.3. In **P5** respondents select their preferred route from each choice set. If the trip diary contains too few trips, or no suitable alternative routes can be generated for a trip, predefined fallback choice sets are used for the remaining questions.

2.2 MyTrips Technical Survey Procedure

The following technical components together form the MyTrips survey tool:

- Conventional online survey (**P1**, **P4**, **P5**): survey parts filled in by respondents

Figure 1: Sequence diagram of the interaction between respondents and the components of the MyTrips survey tool

- Trip diary (**P2**): user interface where respondents enter their trips
- MyTrips web service (**P3**): calculates choice sets and persists survey results

The detailed survey procedure and interaction between these components is shown in Figure 1. The survey procedure is sequential with one exception. After respondents submit the diary two things happen in parallel: respondents continue with the main questionnaire and also the choice sets with alternative routes are calculated. The questionnaire therefore needs to be long enough to allow time for the calculation to finish. For **P3** Choice set calculation an intermodal router is required. For the two case studies we used our proprietary intermodal routing framework Ariadne [18].

2.3 Creation of Alternative Routes for Choice Sets

To calculate SP-off-RP route alternatives, the trip diary is split into trip at all activities except bringing or fetching people and changes between modes of transport. This means that trips are not limited to single modes but can be intermodal. For each trip a set of alternative routes for the SP as well as the RP trips is calculated, using combinations of the available modes of transport for the respondent. The exact method and routing process are survey-dependent. For RP and the SP routes different modes of transport or road and public transport networks can be used.

Once all possible intermodal alternative routes are calculated, choice sets for the SP questions are created. For the two case studies SP-off-RP choice sets in the

▼ Please enter all trips for this day

My Trips

EDIT LOCATIONS

Start: Berggasse 1, Marchegg

Activity: home

+

Mode of Transport: bicycle

☐ shared

From: 06:10

Until: 06:15

+

Place (change): train station Marchegg

Activity: change mode of transport

Change Duration: 5min

+

Mode of Transport: train

line: REX

From: 06:20

Until: 06:40

+

Place (change): train station Gänserndorf

Activity: change mode of transport

Change Duration: 0min

+

Mode of Transport: walk

type: walk

From: 06:40

Until: 06:50

+

Place: Hauptstraße 10, Gänserndorf

Activity: work & education

Duration: 9h 10min

+

Mode of Transport: car

☐ shared ☒ as passenger

From: 16:00

Until: 16:25

+

Place: Berggasse 1, Marchegg

Activity: home

+

SUBMIT DIARY

MyTrips questionnaire contain only two routes. The decision for pairwise choices was taken since the presentation of choice sets was implemented with a map view of and text describing the choices. The restriction to two choices provides enough space for the presentation of both alternatives (see Figure 3).

To avoid SP-off-RP choice sets with unrealistic route alternatives for each trip a subset of feasible routes is selected. Routes that are not viable for most users are discarded since they add no information to the model estimation process. Examples for discarded routes are those with very long walking or cycling stretches or routes that take more than twice as long as the RP route.

The design process for the SP-off-RP alternatives follows the methodology from Arellana et al. [12]. Due to the intermediate routing step the design process needs to be adjusted slightly since the choice sets cannot be directly taken from the calculated design. The adjusted design procedure consists of the following steps.

Step 1 Preliminary design feature definition: For the two case studies mode choice models were based on travel times per mode, waiting time and number of transfers. Due to the difficulty to control the number of transfers in the routing process, the design of the SP alternatives was based only on time variables. Number of changes was only applied in the subsequent model estimation.

Step 2 Optimisation of generic SP design: A 20 row design was chosen as generic design. The generic design was selected using a Bayesian efficient design. The Bayesian D-criterion D_b is calculated as

$$D_b = \int_{\mathbb{R}^k} \{ \det(I)^{-1}(X, \beta) \}^{\frac{1}{k}} \pi(\beta) d\beta, \quad (1)$$

where $I^{-1}(X, \beta)$ is the asymptotic covariance matrix (ACV) of the parameter vector β of k choice model with prior parameter distribution $\pi(\beta)$. Before data is collected, no prior information is available and a normal distribution $N(-2, 1)$ is assumed as the prior for all parameters. Once data is available, a mode choice model with time variables is estimated and the ACV is calculated. For the estimation of D_b a genetic algorithm (GA) was applied in the statistical computing environment R. As Input to the GA a set S of 500 parameter vectors $\beta_s, s = 1, \dots, 500$ was drawn from the prior distribution. For each design in the population of the GA, the ACV was calculated for all β_s to simulate D_b . Once collected data becomes available it is added to the design to assure that the SP choice sets add information to existing data.

Step 3 Selection of alternatives from the generic design: Since the GA is not fast enough to run in real time, SP routes are based on the generic design. Design routes are calculated from the generic design as the percentage of time spent in each mode of transport or for waiting and come in complementary pairs defining a desired choice situation. Ideally the generated choice sets exactly match one of the choice situations.

A reduced generic design example can be seen in Table 1. Route 1A is a unimodal car route and the desired complementary route 1B consists of public transport with walking and waiting.

First, a route from the generic design is matched to the RP route. This is done by finding the design routes with the least differences in the used modes of transport. For a unimodal car route the design route 1A from Table 1 would be the perfect match. For a public transport route including a walking part both 1B and 2A would be perfect matches. In the case of several matching design routes the match is determined by minimising the Euclidean distance between the routes and design route. If the score is still the same for several design routes one of them is chosen randomly from those routes. After matching a design route to the RP route the complementary design route is used to find the best alternative route for this choice set using a similar method as before: first find alternative routes with the best match regarding the used modes of transport and then select the route with the match with minimal Euclidean distance.

Tables

Table 1: First two design route pairs from the generic design.

ID	tt.foot	tt.bicycle	tt.car	tt.PT	Waiting.time
1A	0.0	0.0	1.0	0.0	0.0
1B	0.2	0.0	0.0	0.7	0.1
2A	0.5	0.0	0.0	0.4	0.1
2B	0.0	0.3	0.0	0.6	0.1

2.4 MyTrips SP-off-RP data analysis and mode-choice modelling

In a first step, the SP-off-RP choices are compared to pure SP questions to see if there is a difference between purely hypothetical scenarios ("Would you use the new mode for trips to work?") and SP-off-RP scenarios. This analysis allows a first assessment of whether the hypothetical bias changes between the two settings.

For a more detailed analysis of the choice behaviour different variants of SP-off-RP choice models can be estimated. In this paper Latent Class Models (LCM) were chosen. LCM can be applied to analyse individual heterogeneity (see e. g. [15]). The parametrisation of the class membership model sets them apart from other approaches like Mixed Logit Models. This allows to study how different parts of the population (based on different socio-demographic data) behave in certain mode-choice situations. The MyTrips survey tool collects SP-off-RP data as well as user specific data that can be used in the class membership models. The SP-off-RP data collected for this paper consisted of a mobility diary and for the first three trips one corresponding SP choice each. A non-choice option was also offered to the respondents. The LCM approach assumes that the behaviour of individual n depends on observable utilities and that there is a latent heterogeneity that is unobserved. The latent class approach follows the approach from [15]. This approach is adapted to the SP-off-RP data available here. Assuming respondent n belongs to

class q , the utilities for respondent n to choose alternative j in choice situation t for the RP data is given as

$$U_{jt|q}^{RP}(x_j^{RP}, \epsilon) = \beta_q x_{jn}^{RP} + \epsilon_{jn}, \quad (2)$$

where ϵ_{jn} is independent and identically distributed (*iid*) extreme value with unit scale, x_{jn}^{RP} are the decision variables of alternative j of the RP data for respondent n and β_q are the parameters of the choice model for class q . Assuming the independence of the random unobserved utilities might be a bit of a stretch. To shape the situation in such a way that the *iid* assumption is not too unrealistic the choice set for each trip is chosen as realistically as possible. For this, modes in the choice set are restricted by mode availability, e. g. if a person does not have a driving license a car alternative is not offered or if a trip does not start at home and bike was chosen for the previous trip, a car trip is not offered as an alternative assuming that the car is at home. This allows us to estimate the RP choice as a standard logit. The utility for the SP choice is given as

$$U_{jn|q}^{SP}(x_{jn}^i, \epsilon_j, \eta_j) = \lambda(\beta_q x_{jn}^i + \epsilon_{jn}) + \eta_{jn} =: \lambda(W_{jn|q}) + \eta_{jn}, \quad (3)$$

where x_{jn}^i are the attributes of the SP alternative constructed from the chosen RP alternative i , ϵ_j is the random utility of the RP alternative j and η_j is *iid* extreme value with unit scale and λ is the scaling factor that is needed due to the fact that there might be different scales for the different random variables. The probability of alternative k being chosen by respondent n in the SP experiment conditional on the alternative i being chosen in the RP choice is

$$\begin{aligned} P_{kn|i} &= \text{Prob} \left[\lambda(W_{kn|q}) + \eta_{kn} > \lambda(W_{jn|q}) + \eta_{jn} \forall j \neq k \mid U_{it|q}^{RP} > U_{jt|q}^{RP} \forall j \neq i \right] \\ &= \int \frac{\exp(\lambda(\beta_q x_{kn}^i + \epsilon_{kn}))}{\sum_j \exp(\lambda(\beta_q x_{jn}^i + \epsilon_{jn}))} f(\epsilon \mid U_{it|q}^{RP} > U_{jt|q}^{RP} \forall j \neq i) d\epsilon. \end{aligned} \quad (4)$$

Draws for conditional densities can be constructed from draws from a uniform between zero and one according to the methodology given in [11]. The probability of respondent n choosing k in the SP choice situation and i in the RP choice situation is given by

$$P_{kin|q} = P_{kn|i} \frac{\exp(\beta_q x_{in}^{RP})}{\sum_j \exp(\beta_q x_{jn}^{RP})}. \quad (5)$$

In case there are less than three trips in the RP set of a respondent the probability of the SP choice of a pure SP question is then just a normal logit probability. Equally, if the SP problem produces a no choice, the probability reduces to the logit RP choice.

The class membership probability for class $q \in Q$ of respondent n is given by

$$H_{nq} = \frac{\exp(\theta_q z_n)}{\sum_{r \in Q} \exp(\theta_r z_n)}, \quad (6)$$

where z_n are the respondent's socio demographic variables and θ_q are the parameters of the utility model of class q . The overall SP-off-RP probability for choice situation t of respondent n is then given by

$$P_{tn} = \sum_Q H_{nq} P_{kin|q}. \quad (7)$$

Using the draws from the conditional densities above, it is then straight forward to estimate the model by maximising the negative simulated log likelihood SLL as

$$SLL = - \sum_N \ln P_{tn} = \sum_N \ln \left[\sum_Q H_{nq} \left(\prod_{t \in T_n} \tilde{P}_{k_t|i_t} \right) \right], \quad (8)$$

where N is the set of respondents, T_n is the set of SP-off-RP choice situations of respondent n and $\tilde{P}_{k_t|i_t}$ is the simulated probability of respondent n choosing k_t and i_t in the SP and RP choice situations respectively.

3 Scenarios of the Case Studies

The MyTrips survey tool was already successfully applied to two real-world case studies. The first of these studies studied the extension of an urban subway line (Subsection 3.1) the second the overhaul of the public transport system in a rural setting together with the introduction of micro transit (Subsection 3.2). The following subsections present the settings of these case studies as well as the necessary adaptations to the routing tool and the data collection methodology.

3.1 Case Study 1: Urban Subway Extension

In February 2018 a survey with adults living in the region south and south-east of Vienna was conducted to explore the impact of the extension of the Viennese subway line U1 by 5 stations together with the accompanying changes to the surrounding bus and tram lines. 550 respondents completed the survey. A requirement for the respondents was that they regularly travel through a specified area of interest on work days. The study area was adjacent to the subway extension where the public transport system was adjusted. Since the survey took place after the subway extension and for RP the previous mobility behavior is required participants were instructed to fill in the trip diary for a typical work day before the extension where they travelled through the study area.

OpenStreetMap data from before and after the subway extension were used for routing both individual and public transport. Since OpenStreetMap does not contain timetables they had to be estimated by assuming a standard interval for each line extracted from public transport timetables available in PDF format.

The creation of SP-off-RP questions was adapted as follows. For each user at least three choice sets were shown. If less than three choice sets could be calculated, the SP-off-RP questions were complemented by a set of pre-calculated choice sets where the new subway extension was used in one of the routes of the choice set. Available modes of transport for the alternative routes were walking, cycling, bike-sharing, driving, car-sharing, public transport and intermodal combinations thereof. In [19]

it was shown that weather influences demand of cycling in mode choice behavior, hence for each choice set a randomly determined weather setting was defined and presented to the respondent: temperature (cold, average, hot) and rain (no, light, heavy). In addition respondents were not forced to choose one of the two choices but an additional “no choice” option was offered, in case none of the offered routes was acceptable. In that case a free text answer to explain why none of the routes was acceptable was also requested.

3.2 Case Study 2: Rural Micro Transit

The first case study did not need a lot of imagination from the respondents since the subway is a well known mode of transport for most people travelling in Vienna. Therefore, respondents can imagine the hypothetical scenario quite well. However, the MyTrips survey tool can also be applied in more complex cases where respondents are less familiar with the hypothetical changes to the transport system.

In July 2018 a second case study with 300 respondents aged 16 and above living in the Austrian part of the Vienna Basin was conducted by face-to-face surveys in the study area. The exact survey area was confined by the Northern Railway in the north, the Leitha Mountains in the east, the Raaber Railway in the south and the Southern Railway in the west.

The public transport network as of 2018, which consists of railways and busses, can hardly compete with motorised individual traffic, because aside from the railway lines the busses operate infrequently and routes often involve long detours to connect most of the small villages in the countryside.

The second case study had a futuristic setting which involved the replacement of all current bus lines with new bus lines connecting to the railway lines with an interval timetable in such a way that the area is roughly covered with a grid of fast high-level public transport lines. In addition, a micro transit system is envisioned as follows: former and current bus stations as well as railway stations also serve as micro transit stations. The micro transit works like stationary car-sharing but uses small novel electric vehicles with a maximum passenger capacity of three adults. They can be driven by adults with a driving license in a fast mode up to 45kph and by people older than 16 years without a driving license in a restricted mode up to 25kph. The latter user group is also only allowed to drive on minor roads. A major novelty of the vehicles is their ability to be coupled so that one driver can operate up to five vehicles at once which opens up possibilities for efficient redistribution logistics such as the redistribution done by users themselves during regular trips. The survey area and the envisioned public transport system can be seen in Figure 4 together with the concept of the engageable micro transit vehicles.

Data used for calculating routes was OpenStreetMap for individual transport and a combination of GTFS Open Government Data for the Viennese public transport and a GTFS export for trains and busses in the rural areas provided for the project “SynArea II” by local transport operators (Verkehrsverbund Ostregion). For the future scenario lines and timetables for public transport were created manually.

The creation of choice sets was adapted as follows. For each user at least three and at most five choice sets were shown. Both routes of a choice set used the transport system of the future scenario. The first route was a reconstruction of the route from

Figure 4: Case study 2: Survey area (highlighted), existing railway lines (blue), new bus lines (green), micro transit stations (green circles), park & ride stations (red diamonds and stars) on the left and concept of the engageable micro transit vehicles on the right. Map © OpenStreetMap contributors

the trip diary, i. e. in case a bicycle, motorcycle, or car was used for parts of a trip, these parts of the route were retained. For public transport trips the routes were recalculated with the new public transport system, i. e. using the new bus lines and micro transit. Depending on the respondents' ability to cycle and drive as well as the availability of a bicycle and a car alternative routes were calculated for the reconstructed route for walking, cycling, driving, public transport, bike & ride, and park & ride. Public transport was assumed to be available for every respondent. Then the second route was determined using the design routes as described above.

In order to construct a plausible set of alternative routes the following restrictions applied to the route calculation: for park & ride only routes using a train were considered because combining a private car with a bus or micro transit was not deemed a viable alternative. In addition, the speed of bicycles and cars was reduced for bike & ride and park & ride routes after a certain distance so that switching to public transport is forced to happen at a station nearby. The same was done for the micro transit vehicles to avoid routes using micro transit vehicles running parallel to high-level bus lines.

To address the issue of incomplete diaries, mistakes in the diaries and aborted questionnaires, in this case interviewers performed the task of filling in the surveys while questioning the respondents. The changes in collection lowered the termination rate and improved RP data quality considerably.

4 Results of the two Case Studies

In this section the results of the data analysis are presented for the two case studies. The first subsection looks at first indications of the reduction of the hypothetical bias and at general results of the SP-off-RP survey. Subsection 4.2 presents the results for a first SP-off-RP LCM for the second case study.

4.1 General Survey Results

In the first case study only around 40% of the participants who started the MyTrips online survey completed it. The rest stopped during the trip diary phase. This abortion rate of 60% is unusually high for online surveys. While the reasons for these abortions is not known, the feedback of 150 participants regarding their experience in the form of grades and free text highlights challenges of the survey procedure. The average rating of 2.3 for the trip diary is good (1 = excellent, 5 = bad). However, a frequent note was that filling in the diary takes much longer than typical surveys and is quite complex.

While in general, a trend to overstate usage or willingness to pay for new goods can be seen in SP surveys, only 4% of the respondents in the subway extension survey stated that the new subway option was chosen over their previous RP mode selection. However, in the answers for pre-calculated routes a much higher preference for the subway extension could be found. This shows that putting the SP choice set in a known context has an effect on the choice of the respondent.

Analysis of second case study showed that when asked how likely the usage of the micro transit vehicles for trips with different purposes would be, the answers for likely and very likely range from 28% that state that they would use micro transit as a collective call taxi for work trips to 53% that stated they would use it for trips late at night (see Figure 5). In the SP-off-RP questions in only 13.52% out of a total of 913 questions a micro transit alternative was possible and chosen. This suggests that the hypothetical bias towards stating the willingness to use a new mode is reduced within the SP-off-RP setting.

Figure 6 shows the differences in travel times between the SP choice sets in which the micro transit option was chosen and in which it was not. The figure shows that when the micro transit route is chosen, the travel times the saved time in mode car is largest, i. e. chosen micro transit routes do replace mostly car routes (on average there are six minutes less travel time in the car per route). Micro transit alternatives that were not chosen are on average about eight minutes longer than the chosen alternative suggesting that time sensitive respondents chose car and bike instead of public transport routes to save time and avoid walking.

4.2 Results of the Latent Class Analysis for Rural Micro Transit

LCMs can be applied to study mode choice behaviour for different parts of the population in more detail. In this paper we look at the influence of travel times and the number of changes on the mode choice of classes where class membership is influenced by personal and household variables. The results of the model estimation can be seen in Table 2.

The estimation results show that the class model mostly depends on the availability of public transport options (for accessibility of public transport, respondents were asked to state how easy it is for them to reach public transport with answers in a five level Likert scale), the availability of a travel pass and of a motorbike license. Household variables like household size, employment status and age were not significantly different from 0 at a 10% level. Public transport variables make it more likely that the respondent belongs to class 2, the motorbike license makes it more likely that someone belongs to class 1. The alternative specific constant of class 2

Table 2: Parameter values, t-values, p-values and significance levels of the LCM parameters. Significance codes are: '***': 0.001, '**': 0.01, '*': 0.05, '●': 0.1.

Parameter names	parameters	t-value	p-value	significance level
$\beta_{c1}^{changes}$	0.67	7.87	0.00	***
$\beta_{c1}^{tt foot}$	-5.14	-247.73	0.00	***
$\beta_{c1}^{tt car}$	2.43	11.48	0.00	***
$\beta_{c1}^{tt bike}$	-0.69	-4.77	0.00	***
$\beta_{c1}^{tt PT}$	-0.41	-1.36	0.17	
$\beta_{c1}^{tt micro PT}$	0.06	0.54	0.59	
$\beta_{c2}^{changes}$	0.05	0.54	0.59	
$\beta_{c2}^{tt foot}$	-1.25	-6.50	0.00	***
$\beta_{c2}^{tt car}$	1.63	12.00	0.00	***
$\beta_{c2}^{tt bike}$	-7.23	-68.86	0.00	***
$\beta_{c2}^{tt PT}$	2.55	12.99	0.00	***
$\beta_{c2}^{tt micro PT}$	-0.40	-1.90	0.06	●
θ_{c2}^{ASC}	-1.77	-2.40	0.02	*
$\theta_{c2}^{age > 35 \& < 65}$	-0.42	-1.18	0.24	
$\theta_{c2}^{partially employed}$	0.18	0.45	0.65	
$\theta_{c2}^{pension}$	0.73	1.35	0.18	
$\theta_{c2}^{two pers hh}$	0.33	0.63	0.53	
$\theta_{c2}^{three pers hh}$	0.46	0.77	0.44	
$\theta_{c2}^{four+ pers hh}$	0.94	1.60	0.11	
$\theta_{c2}^{license mbike}$	-1.10	-2.11	0.04	*
$\theta_{c2}^{PT travel pass}$	0.81	1.72	0.09	●
$\theta_{c2}^{PT accessible}$	0.78	1.85	0.06	●
$\theta_{c2}^{PT possible}$	-0.52	-1.46	0.15	
λ	-0.14	-11.61	0.00	***

shifts the probability towards a membership in class 1. Travel time variables show that respondents from class 1 are more sensitive to travel times on foot and prefer individual mobility to public transport (including micro transit) whereas class 2 respondents are more likely to use public transport options.

Overall, the LCM results suggest that household characteristics do not have much influence on mode choice. The possibility to reach public transport and the availability of transport passes shift the modal use towards public transport. The introduction of micro transit improves the accessibility of public transport but the restriction to current bus stops prevents that people who in the previous situation could not reach public transport now have viable micro transit options. In addition, the model shows that in rural areas there is a strong bias towards using cars over other options.

5 Conclusions and Future Work

In this paper, we presented the MyTrips survey tool for the collection of routing based SP-off-RP data. The choice sets for the stated choice experiment are constructed based on mobility diaries filled in by respondents and a connected routing step. This tool allows the collection of complex SP-off-RP data that is suitable for usage in mode and route choice decision modeling and helps to reduce the hypothetical bias. In addition we present the two case studies that applied the MyTrips survey tool and results of these case studies including a SP-off-RP LCM.

The first MyTrips case study for a subway extension in Vienna showed promising results. While the completion of the online diary was deemed too time-consuming

by some respondents and as a result many diaries were incomplete, the resulting SP-off-RP questions seemed to encourage a less pronounced bias towards the new alternative compared with pure SP cases from literature. In the second case study a micro transit system in rural areas was studied. Again, looking at pure SP questions about the willingness to use micro transit and the SP-off-RP question, the willingness to use micro transit was overestimated in the pure SP setting. For the second case study, a LCM was estimated. The model shows that it is not socio-demographic variables that determine the likelihood of public transport usage versus individual modes.

As an extension to the MyTrips survey tool the integration of automated data collection (see e.g. [20]) into MyTrips is planned, such that SP questions can be asked right after some actual route in the area of interest. This would on the one hand minimise the effort of entering a travel diary and on the other hand allow for asking hypothetical questions about a route while it is still fresh in the mind of the traveler. To allow route choice studies, MyTrips will be adjusted for the collection of the trip diary in a more detailed way such as recording a GPS trajectory.

Abbreviations

ACV: asymptotic covariance matrix, GA: genetic algorithm, GTFS: general transit feed specification, LCM: latent class model, RP: revealed preference, SP: stated preference SP-off-RP: stated preference off revealed preference

Availability of data and materials

The data sets generated and/or analysed during the current study contain personal information and as a result are not publicly available due data protection reasons but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is part of the scientific project "SynArea II" supported by the Austrian Research Promotion Agency under grant 860450.

Author's contributions

Both authors together developed the overall collection methodology and the general survey design. MS was responsible for technical conception, user interface design, and implementation of the MyTrips survey. CR was responsible for statistical methodology, choice set design and data analysis.

Acknowledgements

Data collection for both case studies was carried out by INTEGRAL Marktforschung. The authors would like to thank the project partners AMSD Advanced Mechatronic System Development KG, ÖBB-Personenverkehr AG, komobile w7 GmbH and Technische Universität Graz – Vehicle Safety Institute for their valuable feedback and input.

References

1. Brownstone, D., Small, K.A.: Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice* **39**(4), 279–293 (2005)
2. Hensher, D.A.: Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological* **44**(6), 735–752 (2010)
3. Chiu, C., Guevara, C.A.: Assessment of hypothetical bias in the estimation of the vot using sp and sp-off-rp data. In: *International Choice Modelling Conference 2019* (2019)
4. Fifer, S., Rose, J., Greaves, S.: Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? *Transportation Research Part A: Policy and Practice* **61**, 164–177 (2014)
5. Rose, J.M., Bliemer, M.C.J., Hensher, D.A., Collins, A.T.: Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological* **42**(4), 395–406 (2008)
6. Hensher, D.A., Rose, J.M.: Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. *Transportation Research Part A: Policy and Practice* **41**(5), 428–443 (2007)
7. Hensher, D.A., Rose, J.M., Collins, A.T.: Identifying commuter preferences for existing modes and a proposed Metro in Sydney, Australia with special reference to crowding. *Public Transport* **3**(2), 109–147 (2011)
8. Rose, J.M., Hensher, D.A.: Demand for taxi services: new elasticity evidence. *Transportation* **41**(4), 717–743 (2014)
9. Cascajo, R., Garcia-Martinez, A., Monzon, A.: Stated preference survey for estimating passenger transfer penalties: design and application to madrid. *European Transport Research Review* **9**(3), 42 (2017)

10. Train, K.E., Wilson, W.W.: Monte Carlo analysis of SP-off-RP data. *Journal of Choice Modelling* **2**(1), 101–117 (2009)
11. Train, K., Wilson, W.W.: Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B: Methodological* **42**(3), 191–203 (2008)
12. Arellana, J., Daly, A., Hess, S., de Dios Ortúzar, J., Rizzi, L.I.: Development of Surveys for Study of Departure Time Choice: Two-Stage Approach to Efficient Design. *Transportation Research Record: Journal of the Transportation Research Board* **2303**(1), 9–18 (2012)
13. van Cranenburgh, S., Chorus, C.G., van Wee, B.: Vacation behaviour under high travel cost conditions – A stated preference of revealed preference approach. *Tourism Management* **43**, 105–118 (2014)
14. Yu, B., Zhang, J., Fujiwara, A.: Rebound effects caused by the improvement of vehicle energy efficiency: An analysis based on a SP-off-RP survey. *Transportation Research Part D: Transport and Environment* **24**, 62–68 (2013)
15. Greene, W.H., Hensher, D.A.: A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* **37**(8), 681–698 (2003)
16. Campigotto, P., Rudloff, C., Leodolter, M., Bauer, D.: Personalized and Situation-Aware Multimodal Route Recommendations: The FAVOUR Algorithm. *IEEE Transactions on Intelligent Transportation Systems* **18**(1), 92–102 (2017)
17. Rudloff, C., Leodolter, M.: Personalization of Routing Services: User Classification as a Method to Overcome the Cold Start Problem for Personalized Models. *Transportation Research Record: Journal of the Transportation Research Board* **2666**, 38–46 (2017)
18. Prandtstetter, M., Straub, M., Puchinger, J.: On the way to a multi-modal energy-efficient route. In: *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pp. 4779–4784. IEEE, Vienna, Austria (2013)
19. Rudloff, C., Leodolter, M., Bauer, D., Auer, R., Brög, W., Kehnscherper, K.: Influence of Weather on Transport Demand: Case Study from the Vienna, Austria, Region. *Transportation Research Record: Journal of the Transportation Research Board* **2482**, 110–116 (2015)
20. Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., Maurer, P.: Supporting large-scale travel surveys with smartphones – A practical approach. *Transportation Research Part C: Emerging Technologies* **43**, 212–221 (2014)