Inferring proxy response in household travel surveys with unknown completer using a group-based choice model

Takuya Maruyama¹ · Kenta Hosotani¹ · Tomoki Kawano¹

Published online: 3 October 2019
© The Author(s) 2019

Abstract

A proxy response is often accepted for household travel surveys to reduce the survey cost and increase the sample size, but proxy-response biases may be introduced into the sample data. To investigate and correct the bias, completer information for the survey is important, but such information is not always available in practice. This study proposes a novel model that can be applicable in situations where completer information is unavailable. The method introduces group-decision modeling in analyzing the response choices of the household travel survey, where the survey response is considered to be a task allocation among household members. The proposed model can infer the probability of proxy response and the proxy-response bias of trip-related records without completer information. The potential of the proposed model was confirmed by application to a household travel survey in Japan. The inferred probability of the proxy response and the inferred bias without completer data demonstrated surprisingly similar results to the existing study with actual proxy-response data. Specifically, the model inferred a high probability of proxy response in young adults and a low proxy probability in middle-aged females, and the model inferred the proxy-response bias that female proxy respondents in the middle-aged group report lower trip rates than self-respondents. This method will be valuable not only in travel surveys, but also in the general research and practice of social surveys.

Keywords Proxy-response bias · Completer information · Travel survey · Group-based discrete choice

Introduction

Social surveys, such as travel surveys, are conducted to obtain data for planning public policies, and thus, high quality data is crucial. Recently, survey methods have shifted from interview-based methods to self-reported paper-based and web-based survey methods to reduce survey costs and the burden on investigators. Owing to this change, several biases have been introduced in survey data that must be carefully examined. While many types of

✉ Takuya Maruyama
takumaru@kumamoto-u.ac.jp

¹ Kumamoto University, Kumamoto, Japan

biases exist (e.g., non-response bias and coverage bias), this study focuses on the proxy-response bias in household travel surveys.

Proxy response in household travel surveys can be defined as the situation in which a household member responds (i.e., completer) to a travel survey on behalf of the other members in the household (i.e., proxy respondents). This is in contrast with self-response in which a household member responds to his/her own travel survey. For example, the mother in a household may report the estimated travel data of her entire household, which may include her father, husband, and children. However, her father may visit a hospital, walk to the park, or go to the community center, and the completer (mother) may not be aware of these trips. Therefore, such crucial records are inadvertently omitted, and thus, the reported number of trips for her father is underestimated. This is an example of proxy-response bias, which is obviously problematic. Several authors have investigated proxy-response bias using survey data with a special format to collect the completer information (Badoe and Steuart 2002; Richardson 2006; Wargelin and Kostyniuk 2006). They found trips were often underreported for individuals responding by proxy, which produced substantial bias in the travel survey data, so collecting information about the person responding by proxy is highly important. Stopher (2012) stated that the reporting proxy-response statistics provide an important measure of survey quality. Unfortunately, travel surveys do not always collect information on completers in practice, which causes problems in accurately understanding travel behavior and for comparing household travel survey data among several cities and time periods (e.g., intercity and temporal comparisons). Therefore, the development of a methodology that enables us to infer the proxy response and correct the proxy-response bias in the situations where completer information is unavailable is highly important. This is the target of the present study.

Choosing the completer of the requested household travel survey can be regarded as task allocation of household activities among household members. Many studies have developed household task-allocation models considering group-based decision choice in the transportation research field. This study proposes a method that is similar to one of these task allocation models. This model can infer the probability of proxy response for each member in household, and infer the proxy-response bias. Such inference is possible without collecting information on the completers (i.e. with unknown completer). Therefore, the objectives of this study are:

- to develop a group-based response choice model for household travel surveys,
- to retrieve the probability of proxy response from the group-based choice model, and
- to infer the proxy-response probability and proxy-response bias from real household travel survey data with unknown completer using the developed model.

Although technological innovation has provided new types of travel data (e.g., global positioning system (GPS) tracking data and smartcard data), these data often lack demographic information. Some of these data focus on single travel modes. Therefore, these new data can complement household travel surveys, but they are not sufficient to replace household travel surveys in the short term (Verreault and Morency 2018). In addition, collecting attitudinal answers for transport policies should be based on social surveys. Therefore, examining methods to improve classic household travel survey data are still important.

Contributions of our study are summarized as follows. First, we proposed a novel method to infer the proxy-response probability of household travel surveys with unknown completer using a group-based discrete choice model. This innovation was made possible by incorporating group-decision modeling, which is popular in activity-based travel

analysis, in the study of the travel survey method. Our proposed method was equivalent to a type of household-based model suggested by Zhang et al. (2009), but we provided another interpretation of their model. Our methodological innovation will be highly useful in fields other than transportation research. Second, using the proposed model, we can more clearly describe the response choice behavior of the household travel survey. This description will be useful for considering strategies to increase the response rates and to reduce survey bias, in general. Third, we inferred the proxy-response bias using the proposed method and found that the inferred bias was consistent with the proxy-response bias reported in the literature. This finding was surprising because our proposed method made the inference without completer data and showed similar trends to existing studies with real completer data, thus indicating the validity of our model.

The remainder of this paper is organized as follows. "Second section" reviews existing literature on proxy response and the group-based decision choice model. "Third section" describes the methodologies used in this study. "Fourth section" demonstrates the results and discussion. Finally, "Fifth section" concludes the study.

Literature review

Proxy response in travel surveys

Problems of proxy-response bias have been examined in many fields, including survey research, epidemiology, health behavior, mental health studies, and marketing. Moore (1988) and Cobb (2018a, b) provide comprehensive reviews on this topic. This subsection provides a brief review of proxy-response bias in transportation research, and "[Proxy response in general surveys](#)" section reviews the research on proxy-response bias in other fields.

Badoe and Steuart (2002) investigated proxy response in a telephone-interview-based travel survey conducted in the Greater Toronto Area, Canada. Home-based discretionary trips and non-home-based trips were underreported by proxy-response, especially short- and medium-distance trips by car. The number of available vehicles in the household, gender, and driver's license status were related to the underreporting of trips. Badoe and Steuart (2002) further demonstrated adjustment factors to correct the proxy-response bias.

Wargelin and Kostyniuk (2006) investigated proxy-response bias in the National Household Travel Survey (NHTS) conducted in four U.S. cities by the computer-aided telephone interview method. The proxy respondents tended to include younger male adults, people with less education, member of a larger household, or member in a high-income household. Proxy respondents reported a significantly lower trip rates than self-respondents, and no trips were often reported for the proxy respondents. Wargelin and Kostyniuk (2006) concluded that the largest bias might occur for young and middle-aged adult females.

Bose and Giesbrecht (2004) also reported that in the 2001 NHTS, proxy respondents included males between 18 and 24 years in age, non-drivers, people with less education, people who were absent from home on the travel day, or disabled individuals. Trip rates of proxy respondents were much lower than those of self-respondents. Lower trip rates were also reported for proxy respondents for annual long-distance trips, weekly walking and bike trips, and trips by public transportation.

Richardson (2006) examined proxy-response bias in travel diary surveys in the Coastal South-East Queensland, Australia. Young adults and males were more likely to have

proxy responses. Females, especially those between 25 and 45-years-old, tended to self-respond and complete a proxy-response for other household members. Proxy respondents were found to have made fewer trips, shorter travel times, and shorter travel distances than self-respondents.

Chung et al. (2017) developed a web-based travel survey system with the intention of reducing proxy response. In their case study, 27% of respondents were proxy or partial-proxy respondents, which was almost half that of the computer-assisted telephone interviewing method. As found in other studies, trips were underreported for proxy respondents compared to self-respondents.

Ashley et al. (2009) demonstrated that the proxy response of travel diary entries tended to be completed on the days after traveling, indicating a delay in completion, more frequently than self-respondents. Verreault and Morency (2015) examined the change of proxy-response bias over time using five phone-based household travels surveys conducted in the Greater Montreal Area, Canada between 1987 and 2008, demonstrating that the proportion of self-respondents increased over time, partly due to the decrease in household sizes. However, self-respondents aged 16–30 have decreased over time. Verreault and Morency (2015) concluded that proxy-response bias is decreasing and this is partly caused by decreases in household sizes and trip rates.

In summary, proxy-respondents are likely to be male and young adults. Proxy respondents tend to have lower trip rates than self-respondents. These findings based on real proxy-respondent data are compared to our results in "[Results and discussion](#)" section.

Proxy response in general surveys

In the field of environmental behavior, Seebauer et al. (2017) revealed the proxy-response bias of household energy conservation and pro-environmental values. In their examination of travel behavior, bias was higher for shopping or leisure trips than work trips. Reynolds and Wenger (2012) examined proxy response in the U.S. Current Population Survey. Half of the labor force data in the survey was given by proxy-response. Reynolds and Wenger (2012) found that self-reported wages were higher than proxy-reported wages. Moreover, proxy response produced bias in gender wage-gap estimates because the gender composition of proxy respondents has changed remarkably during the last 30 years due to the increase in the female labor force. Only 28% of female responses were provided by proxy in labor market information in 1979, which increased to 40% in 2009.

Several modeling approaches were used to examine this issue. For example, Wagmiller (2009) applied the fixed-effect regression model to assess the response bias in adult children reporting their parents' education level. In health studies, Skolarus et al. (2010) evaluated the difference between stroke patient self-response and proxy response using linear regression models and simulated proxy biases. Todorov and Kirchner (2000) demonstrated proxy-response biases in the National Health Interview Survey on Disability, and Todorov (2003) developed regression models to predict self- and proxy reporting and to predict the differences between self- and proxy responses.

The propensity score method has been widely used in medical research to investigate survey bias. A propensity score is the estimated probability of a given individual to respond by proxy. This method is generally used for bias reduction in comparisons between treatment and non-randomized control groups. Observational data is limited because the allocation of self-response and proxy response is nonrandomized, which leads to selection

bias. Therefore, the propensity score method provided an approximation to the survey bias in a randomized controlled experiment of proxy and self-response.

Ellis et al. (2003) used the propensity score method of matched sampling to compare proxy and self-response in the Medicare Health Outcome Survey. Logistic regression (or binary-logit model) was used to calculate the probability that a given individual belonged to the proxy-respondent group. Elliott et al. (2008) assessed the proxy-respondent effect on health care evaluation by Medicare beneficiaries using propensity-score-weighted regression. Li et al. (2015) compared the proxy- and self-reported results in health and functional status measures using the Medicare Current Beneficiary Survey by propensity score matching.

All existing studies used survey data that included a question about the individual who completed the survey. Therefore, these studies were possible because of the information on the self- and proxy respondents. However, the proposed methodology will be valuable because we can evaluate these issues, even if we do not have such information. In addition, these existing studies calculated the propensity score using the individual-based binary-logit model. Our proposed group-based choice model is expected to refine the choice probability calculation because our model describes the task allocation of survey response in the household, which is expected to be a better representation of the response choice behavior.

Group-based decision model

Survey participation can be regarded as task allocation among household members. Here, we consider the survey response to be a task or activity, and the choice of who will complete the survey can be considered an allocation of activities among household members. A growing number of studies on household task allocation modeling can be found in transportation literature. Ho and Mulley (2015) review intra-household interaction modeling in transportation research, and de Palma et al. (2014) review group decision-making models, not only in transportation, but in labor economics and family economics. There are two special issues published on intra-household interactions and group decision-making (Bhat and Pendyala 2005; Timmermans and Zhang 2009). Here, we review the studies relevant to our research target.

Srinivasan and Athuru (2005) examined the allocation of out-of-home maintenance activities among household members by proposing a nested and joint mixed-logit modeling framework to consider household person allocation of activities. The results suggested that life-cycle and household role, income, gender, and employment status affect the person allocation decision. Srinivasan and Bhat (2005) developed a joint mixed-logit hazard-duration (discrete-continuous) model and applied the model to out-of-home maintenance activity generation. The choice structure assumed that shopping was either allocated to one of the household heads or to both. In other words, four alternatives were assumed: no shopping, male shopping, female shopping, and joint shopping. Srinivasan and Bhat (2006) developed a multiple discrete-continuous model of joint discretionary activity allocation and duration for a couple in a nuclear family. The activity allocation involved five discrete-continuous choices: (1) male's independent in-home choices, (2) female's independent in-home choices, (3) male's independent out-of-home choices, (4) female's independent out-of-home choices, (5) household's joint out-of-home choices. Age, employment status, student status, and personal vehicle availability, among other

individual characteristics, impacted discretionary-activity allocation decisions. Bhat et al. (2013) formulated and estimated a multiple discrete-continuous extreme value (MDCEV) model of household-level activity pattern generation that predicted the independent and joint activity participation decisions of all members in a household for all combinations of members.

Zhang et al. (2009) summarized several types of household utility functions to represent the group decision-making of household behavior, such as multi-linear household utility function and iso-elastic household utility function, among others. Using the multilinear utility function approach, Ermagun and Levinson (2016) developed a group decision-making model for school trip accompaniment of children. Weiss and Habib (2018), however, criticized the multi-linear utility function approach in existing group-decision models by stating that multilinear utility function is not suitable for discrete choice modeling using random utility maximization. Instead, Weiss and Habib (2018) proposed a generalized parallel constrained choice model (Gliebe and Koppelman 2005) for analyzing the intra-household escort of school students. In the original parallel constrained choice model (Gliebe and Koppelman 2005), the group utility function was derived from the generalized extreme value (GEV) model to capture tour patterns. Ho et al. (2016) proposed a similar group-based model derived from the GEV model to analyze the value of travel time savings.

Zhang et al. (2009) demonstrated that the maximum-type household utility function was a special case of the iso-elastic household utility function. The maximum-type utility function can be approximated by the logsum variable that makes the model operational. The resultant model was denoted the maximum utility (MAX) model. Note that our approach is equivalent to a type of the MAX model by Zhang et al. (2009); however, we demonstrate a different model derivation and interpretation of the MAX model.

Methodology

Assumptions

Proxy-response problems can be analyzed in various situations of available data. If the data has information on the actual completer, we can investigate the proxy respondents. Unfortunately, this information is not always available in social surveys targeting household-based responses. As mentioned earlier, we focus our study on the situation in which survey completer data is unavailable. In these cases, we need another data set in order to be able to infer the proxy response. The data set required is the information on non-responding households. Certainly, this data may not be available either in most surveys, but, if the survey is organized by the government, the survey organizer should have some information on the target households. For example, the data in the present study includes basic information of non-responding households, namely the household size, the gender and age (in 10-year units) of each household member, and the zone of their residence.

We assume that one household member completes the survey for all members in the household. We call this member the *representative completer*. In reality, there can be more than one survey completer in the household, as reported in Wargelin and Kostyniuk (2006) and Richardson (2006). We adopt the assumption of one survey completer for simplicity, and this assumption enables us to examine the case that survey completer data is unavailable. The multi-completer situation could be a promising future study if survey completer data is available. In summary, we have the following assumptions in our analysis.

- One member in the household (representative completer) responds to the survey for all other members in the household.
- Information about the representative completer is unavailable.
- Basic information on the members of non-responding households is available.

Model

We assume that household i has J_i members and one-member j is a representative completer. The probability that household i responds through representative completer j is defined as $P_{i,j}$. The probability that household i does not respond is $P_{i,0}$. If the utility function for “no-response” is defined by $V_{i,0}$ and that for response by member j is $V_{i,j}$, the choice probability of each alternative can be expressed by the following multinomial logit model with the well-known assumption of error term (independent and identically distributed Gumbel distribution) for utility functions:

$$\begin{aligned} P_{i,0} &= \frac{\exp V_{i,0}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}} \\ P_{i,j} &= \frac{\exp V_{i,j}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}} \quad \text{for } j \neq 0 \end{aligned} \quad (1)$$

Note that the choice set of this model is variable across households. The size of the choice set is $J_i + 1$ for household i ; that choice set includes the situation of “no-response” and that in which member j in J_i will respond as the representative completer. If the data for the representative completer is available, Eq. (1) can be a standard multinomial logit model, and the standard estimation procedure will be possible. However, we assume that the data for the representative completer is unavailable, and thus, this model will not be the same as a standard multinomial logit model.

We can observe the data of household-based non-response or response, which correspond to probability $P_{i,0}$ and $\sum_j P_{i,j}$, respectively. Then, we can estimate the model with this partial observation. We maximize the likelihood L^* (or the log likelihood L) to estimate the parameters in the model as follows:

$$L^* = \prod_{i \in N} P_{i,0}^{1-\delta_i} \left(\sum_{j=1}^{J_i} P_{i,j} \right)^{\delta_i} \quad (2)$$

$$\begin{aligned} L = \ln L^* &= \sum_{i \in N} \left\{ (1 - \delta_i) \ln P_{i,0} + \delta_i \ln \sum_{j=1}^{J_i} P_{i,j} \right\} \\ \delta_i &= \begin{cases} 1 & : \text{household } i \text{ responds} \\ 0 & : \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

This model is equivalent to a type of MAX model in the group-decision models proposed by Zhang et al. (2009). For details of this equivalence, see the “Appendix A1”. “Appendix A1” also discusses how the binary choice indicator δ_i is sufficient to estimate the multinomial choice model in Eq. (1), which usually requires an observed choice indicator to identify the completer.

We can now examine the proxy-response probability using this model. Let A be the event that individual j in household i is the representative completer, and let B be the event that household i responds. The probability of individual j in household i to be a proxy respondent, P_{proxy} , can be theoretically retrieved. Here, we omit indices i and j in A , B , and P_{proxy} to simplify the description. First, the probability that household i responds, $P(B)$, and the probability that individual i is the representative completer, $P(A)$, is given by

$$P(B) = \sum_{j=1}^{J_i} P_{ij}, P(A) = P_{ij}. \quad (4)$$

Then, the joint probability that someone in household i responds and individual j is not the representative completer is given as

$$P(\bar{A} \cap B) = P(B) - P(A) = \sum_{j=1}^{J_i} P_{ij} - P_{ij}. \quad (5)$$

Finally, we have the following equation of the proxy-response probability P_{proxy} , which is the conditional probability that individual j is not the representative completer under the condition that household i responds:

$$\begin{aligned} P_{proxy} &= P(\bar{A}|B) = \frac{P(\bar{A} \cap B)}{P(B)} = \frac{\sum_{j=1}^{J_i} P_{ij} - P_{ij}}{\sum_{j=1}^{J_i} P_{ij}} = 1 - \frac{P_{ij}}{\sum_{j=1}^{J_i} P_{ij}} \\ &= 1 - \frac{\exp V_{ij}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{ij}} / \frac{\sum_{j=1}^{J_i} \exp V_{ij}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{ij}} = 1 - \frac{\exp V_{ij}}{\sum_{j=1}^{J_i} \exp V_{ij}} \end{aligned} \quad (6)$$

Using the above equation, we can infer the probability of proxy response and analyze the bias caused by proxy response. Note that this simple formula is constructed under the assumption that each household has only one completer. If we relax this assumption and allow multiple completers, we will require a more complicated formula for determining the probability of proxy response. In addition, if the household size $J_i = 1$, that is, a single household, then $P_{proxy} = 0$ using Eq. (6). If the household size J_i increases with similar utility functions, the probability of proxy response will increase by Eq. (6). These properties appear to be natural in analyzing the proxy response.

Results and discussion

Data

We used data from a household travel survey, which is referred to as the Kumamoto Person Trip (PT) survey, conducted in 2012 in Kumamoto, Japan. The area covered by the Kumamoto PT survey is the Kumamoto metropolitan area, which includes Kumamoto city and its surrounding areas. In this study, we use a subset of the target sample to estimate the model: 13,279 households (37,232 people). Among this subset, 3752 households (11,438 people) responded to the Kumamoto PT survey, and 9527 households (25,794 people) did

not respond (Maruyama 2014). The inferred probability determined by the proposed model is calculated for the entire target sample.

Group-based response choice

We estimated the proposed model using response and non-response data of the Kumamoto PT survey. We set the following utility functions:

$$V_{i,0} = - \sum_k \alpha_k x_{k,i} \quad (7)$$

$$V_{i,j} = \sum_l \beta_l y_{l,i,j} \quad (8)$$

where α_k , β_l are parameters to be estimated. The household-based observable variables $x_{k,i}$ are introduced in the utility function of non-response $V_{i,0}$ in Eq. (7). The individual-based observable variables $y_{l,i,j}$ are introduced in the utility function of response $V_{i,j}$ in Eq. (8). We introduce the negative sign in Eq. (7) to make the estimated parameter easy to interpret. Estimated parameters α_k and β_l both increase with response tendency in this setting. If we do not introduce the negative sign, the two parameters would have opposite signs, even in the same tendency.

Then, we introduce the following variables in $x_{k,i}$: dummy variable of single household, household size, and dummy variable of land use regulation in the residential zone. Information on the land use regulation is introduced as a proxy variable of the built environment in the residential zone. We use variables of the gender and age of the household member in $y_{l,i,j}$. We cannot use other variables because our method requires the information of the members of non-responding households. Thus, only the basic information of gender and age is available.

Table 1 shows estimation results of the proposed group-based response choice model. The results indicate that response choice behavior was mainly affected by household size and built environment, as well as gender and age of the household members. Focusing on the land use regulations in Table 1, the parameters were significantly positive in the low- and medium-rise exclusive residential districts. This means that the residents in these districts tended to respond to the PT survey. Specifically, the parameters of low-rise exclusive residential districts were more than twice that of medium-rise exclusive residential districts. The precise interpretation of this result requires understanding of the City Planning Law in Japan, and Maruyama (2014) provide an illustration of the response rate distribution and land use regulation. Here, we only provide intuitive interpretations. The reasons for the high response by residents in low-rise exclusive residential districts could be that they live in detached houses, are wealthy, and have interest in public policies in their area. In addition, as shown in Table 1, most parameters of gender and age are significantly positive, and the parameters tend to increase as age increases. Hence, households with elderly adults generally tended to respond to the Kumamoto PT survey. In terms of gender, for young and middle age adults, females tended to respond to the survey more than males, but those aged 70 and over, the tendency of male response was higher.

Maruyama (2014) estimated the individual-based response choice model for the same data of the Kumamoto PT survey, and the adjusted ρ^2 of his model was 0.126 (See "Appendix A2" for detail). Our model reported an adjusted ρ^2 of 0.190. Although direct comparison of his model and our model is meaningless because the model outputs are different, the

Table 1 Estimation results of response choice behavior using a group-based model

Explanatory variables	Parameters	t-stat.
<i>Household characteristics (α_k)</i>		
Constant	-2.62	-9.57***
Single household	0.04	0.47
Household size	-0.22	-10.98***
<i>Land use regulation in residential zone</i>		
Low-rise exclusive residential districts	0.56	2.62***
Medium-rise exclusive residential districts	0.23	3.12***
Commercial districts	0.06	0.61
Quasi-industrial districts	0.22	1.52
<i>Individual characteristics (β_l)</i>		
Male aged 20–29 years	0.09	0.32
Male aged 30–39 years	0.79	2.84***
Male aged 40–49 years	1.17	4.19***
Male aged 50–59 years	1.67	5.30***
Male aged 60–69 years	2.17	7.37***
Male aged 70+ years	2.24	7.38***
Female aged 20–29 years	1.18	4.50***
Female aged 30–39 years	1.82	7.03***
Female aged 40–49 years	2.21	8.14***
Female aged 50–59 years	2.24	8.26***
Female aged 60–69 years	2.17	7.55***
Female aged 70+ years	1.46	4.53***
Observations		13,279
ρ^2		0.192
Adjusted ρ^2		0.190

***1% significant

increase in the adjusted ρ^2 indicates that our group-based model may describe the response choice behavior better than the individual-based model.

Proxy respondent and proxy-response bias

Using Eq. (6), we calculate and analyze the probability of proxy response. Figure 1 provides the average probability of proxy response as a function of age and gender of the respondent. Young males and females (< 19 years old) were equally likely to be proxy respondents, but the probability of proxy response varied remarkably by gender with the increase in age. Working-age males (25–59 years old) were much more likely to respond by proxy than working-age females. This result is surprisingly similar to those reported by Richardson (2006), Wargelin and Kostyniuk (2006), and Ashley et al. (2009) using actual data about the proxy respondent. Please note that our data does not include actual proxy-respondent information, so Fig. 1 is based on inference by our proposed method. The consistency in our results with other studies indicates that our proposed method provides a plausible framework for inferring proxy-respondent information.

Fig. 1 Inferred probability of proxy response by age and gender of respondent

To examine the proxy-response bias using our proposed method, we divided the individual-based sample into two classes: the self-response class and the proxy-response class. The self-response class consists of individuals with less than 50% probability of responding by proxy, whereas the proxy-response class consists of individuals with 50% or greater probability of responding by proxy. The difference in trip-related reporting between the two classes can be an inference of the proxy-response bias.

Figure 2 shows the results comparing the trip rate of the self-response class with that of the proxy-response class by gender and age. The statistical test for the difference in trip rates is reported in Table 2. Male self-respondents recorded higher trip rates than male proxy-respondents. The results confirmed statistically significant differences in males aged 20–54, 70 and over. The difference in female trip rates was great for the younger generation. The results confirmed statistically significant differences in females aged 20–49 and 60–89. We found large gaps in females aged 20–49. Females in these generations often engage in childcare activities and make trips to take their children to kindergarten, visit nursing homes, and shop. These trips could be misreported by proxy response, which could be an interpretation of large gaps in the trip rates by females aged 20–49.

Next, we analyzed the proxy-response bias categorized by gender and employment. Figure 3a and Table 3 show the results of comparing male trip rates by proxy and self-response

Fig. 2 Trip rates by self- and proxy response class, gender, and age. *Note* Self-response class consists of individuals with inferred probability of proxy response < 50%; proxy-response class consists of individuals with the probability $\geq$ 50%

Table 2 Statistical test of trip rate differences by self- and proxy response, gender, and age

Age	Male						Female					
	Self-response			Proxy response			Self-response			Proxy response		
	Trip rate	N	S.D.	Trip rate	N	S.D.	Trip rate	N	S.D.	Trip rate	N	S.D.
15–19				2.14	2294	1.08				2.16	2224	1.06
20–24	2.65	163	1.63	2.20	1377	1.42	2.84	308	1.56	2.18	1698	1.21
25–29	2.78	287	1.59	2.32	1746	1.49	2.96	1135	1.78	2.22	1444	1.26
30–34	2.81	267	1.72	2.59	2066	1.67	3.26	1995	1.96	2.39	970	1.46
35–39	2.91	370	1.84	2.68	2740	1.71	3.39	2755	2.04	2.50	912	1.61
40–44	3.01	359	1.78	2.74	2573	1.76	3.10	2994	1.94	2.63	604	1.85
45–49	2.88	320	1.67	2.71	2415	1.64	2.98	2963	1.85	2.65	732	1.60
50–54	2.99	494	1.84	2.79	2511	1.76	2.88	2819	1.70	2.84	1086	1.76
55–59	2.99	577	1.94	2.87	2884	1.76	2.80	2870	1.69	2.80	1325	1.75
60–64	3.16	1980	1.88	3.15	2533	1.93	2.88	3431	1.73	2.72	1800	1.78
65–69	3.19	2276	1.93	3.11	1196	2.07	2.92	2007	1.84	2.81	1994	1.86
70–74	3.12	2194	2.07	2.95	695	2.03	2.80	969	1.83	2.48	2435	1.80
75–79	2.83	1940	1.97	2.55	498	2.07	2.26	921	1.65	2.09	1949	1.74
80–84	2.47	1352	1.85	2.04	392	1.89	1.90	820	1.63	1.47	1282	1.60
85–89	1.85	585	1.63	1.39	212	1.75	1.28	586	1.39	1.11	707	1.44
90+	1.32	152	1.44	0.93	98	1.64	0.80	325	1.27	0.81	431	1.29
Total	2.91	13,316	1.93	2.66	26,230	1.68	2.88	26,898	1.86	2.33	21,593	1.60

*10% significant, **5% significant, ***1% significant

and working situation. On one hand, there was a large difference in the trip rates for middle-aged workers. On the other hand, the difference of trip rates was minimal for elderly workers. Statistically significant differences were confirmed in male workers aged 25–54 and in male non-workers aged 30–39, 55–59, and 65 and over. We observed a decrease in trip rates for male working generations by proxy response. This bias will be partially caused by business trips often being misreported by proxy response. We see large gaps in both the young generation and the elderly generation for male non-workers. This can be explained by the situation that trips by students and walking and recreational trips by the elderly will be difficult to capture by proxy response.

Figure 3b and Table 4 shows the results of comparing female trip rates by proxy and self-response and working situation. We found similar trends of a trip rate decrease with proxy responses in each working situation. Statistically significant differences were confirmed in female workers aged 20–49 and 80–84 and in female non-workers aged 20–49 and 60–89. Particularly, the trip rates of young female non-workers by proxy response were much lower than those by self-response. Females in those generations have many trips for shopping and childcare, and these trips would likely be misreported by proxy response.

From the above analysis, we successfully demonstrated the empirical inference of proxy-response bias. Interestingly, Richardson (2006) reported similar trends to Fig. 2 and Table 2 using actual proxy-response data. Note that our results are provided without actual proxy-response data and are based on inference using our proposed methodology. These similarities suggest that our proposed methodology successfully inferred the proxy-response bias without completer information. These findings are quite useful because this method may enable us to infer and correct the proxy-response bias, even if we do not have completer information, which is often the case with social survey data.

Conclusion and future works

Proxy response is often utilized in social surveys to reduce survey costs and increase the sample size. Considering the bias caused by proxy response, completer information should be collected and recorded. However, such information is sometimes unavailable in practical surveys. This study proposed a novel model that can be applicable in situations where completer information is unavailable. The proposed model utilized a group-decision modeling

Fig. 3 Trip rates by self- and proxy response class, gender, age, and working situation. *Note* Self-response class consists of individuals with inferred probability of proxy response < 50%; proxy-response class consists of individuals with the probability $\geq 50\%$

Table 3 Statistical test of trip rate differences by self- and proxy response and working situation: Male

Age	Working male						Non-working male							
	Self-response			Proxy response			Self-response			Proxy response				
	Trip rate	N	S.D.	Trip rate	N	S.D.	t-stat.	Trip rate	N	S.D.	Trip rate	N	S.D.	t-stat.
15–19				2.15	2252	1.06					1.24	42	1.46	
20–24	2.63	161	1.63	2.33	1226	1.35	2.27**	4.00	2	1.41	1.15	151	1.52	2.83
25–29	2.81	273	1.57	2.44	1552	1.45	3.69***	2.07	14	1.77	1.36	194	1.44	1.47
30–34	2.90	244	1.69	2.68	1920	1.63	1.91*	1.87	23	1.89	1.40	146	1.74	1.13
35–39	2.98	332	1.80	2.75	2593	1.68	2.17**	2.26	38	2.05	1.33	147	1.53	2.63**
40–44	3.15	321	1.73	2.78	2455	1.75	3.58***	1.87	38	1.88	1.84	118	1.69	0.09
45–49	2.95	279	1.64	2.73	2321	1.62	2.12**	2.39	41	1.87	2.10	94	1.98	0.83
50–54	3.07	417	1.78	2.82	2425	1.75	2.70***	2.57	77	2.12	2.06	86	1.96	1.60
55–59	3.06	460	1.87	2.92	2690	1.72	1.57	2.70	117	2.19	2.20	194	2.08	1.99**
60–64	3.13	1349	1.83	3.17	1915	1.90	−0.50	3.22	631	1.99	3.11	618	2.04	0.97
65–69	3.26	1000	1.93	3.42	605	1.97	−1.56	3.14	1276	1.93	2.79	591	2.11	3.46***
70–74	3.08	719	2.13	3.11	272	1.93	−0.22	3.14	1475	2.03	2.85	423	2.09	2.60***
75–79	2.88	543	1.95	2.71	162	2.01	0.93	2.81	1397	1.98	2.47	336	2.10	2.70***
80–84	2.41	358	1.84	2.53	105	1.93	−0.58	2.49	994	1.85	1.86	287	1.85	5.05***
85–89	1.89	151	1.58	1.76	51	1.84	0.45	1.84	434	1.65	1.27	161	1.72	3.58***
90+	1.52	27	1.45	1.00	12	1.60	0.96	1.28	125	1.45	0.92	86	1.67	1.63
Total	2.99	6634	1.87	2.72	22,556	1.62	10.76***	2.83	6682	1.99	2.27	3674	2.05	13.44***

*10% significant, **5% significant, ***1% significant

Table 4 Statistical test of trip rate differences by self- and proxy response and working situation: Female

Age	Working female						Non-working female					
	Self-response			Proxy response			Self-response			Proxy response		
	Trip rate	N	S.D.	Trip rate	N	S.D.	Trip rate	N	S.D.	Trip rate	N	S.D.
15–19				2.18	2187	1.03				1.05	37	1.69
20–24	2.91	255	1.46	2.27	1537	1.14				1.34	161	1.51
25–29	3.11	820	1.66	2.28	1254	1.13	2.55	53	1.95	1.77	190	1.81
30–34	3.37	1278	1.77	2.50	816	1.38	2.56	315	2.02	1.82	154	1.74
35–39	3.48	1820	1.85	2.57	741	1.38	3.06	717	2.25	2.22	171	2.34
40–44	3.17	2071	1.79	2.71	487	1.50	3.22	935	2.35	2.31	117	2.85
45–49	3.07	2171	1.64	2.77	570	1.45	2.95	923	2.24	2.23	162	1.98
50–54	3.02	2046	1.60	2.97	768	1.65	2.75	792	2.30	2.52	318	1.96
55–59	2.89	1794	1.52	2.96	843	1.56	2.51	773	1.89	2.53	482	2.00
60–64	2.95	1570	1.57	2.91	751	1.67	2.66	1076	1.94	2.59	1049	1.84
65–69	2.95	661	1.77	2.88	556	1.82	2.83	1861	1.86	2.78	1438	1.87
70–74	2.66	346	1.90	2.52	529	1.78	2.90	1346	1.88	2.47	1906	1.81
75–79	2.25	289	1.73	2.18	370	1.76	2.87	623	1.79	2.06	1579	1.73
80–84	2.09	240	1.66	1.66	189	1.66	2.26	632	1.61	1.44	1093	1.59
85–89	1.43	123	1.40	1.57	69	1.89	1.83	580	1.61	1.06	638	1.38
90+	1.14	50	1.52	1.18	34	1.77	1.24	463	1.38	0.78	397	1.24
Total	3.05	15,534	1.71	2.49	11,701	1.39	0.74	275	1.21	2.14	9892	1.87
							2.65	11,364	2.03			

*10% significant, **5% significant, ***1% significant

approach in the response choice modeling of a household travel survey. By considering survey response to be task allocation among household members, this activity-based approach can be incorporated in the data examination. The application of the proposed model to real survey data revealed promising findings demonstrating the potential of our method. In summary, we present the following contributions and findings of this research.

- We have developed a novel group-decision choice model that enables us to analyze response choice behavior in household units.
- We applied the proposed model on response choice behavior to the Kumamoto PT survey, a household travel survey in Japan, and demonstrated that the household size, built environment, and household member age and gender affect the household response choice.
- The proposed model provides an inference of the probability of proxy response and the proxy-response bias of trip-related records without completer information.
- The inferred probability of proxy response in the sample data with unknown completer demonstrated surprisingly similar results to existing studies with actual proxy response data: high probability of proxy response for young individuals, and low proxy probability in middle-aged females.
- The inferred proxy-response bias in the data also demonstrated remarkably similar results to existing studies with actual proxy response data: middle-aged females responding by proxy indicate lower trip rates than self-respondents.

Although the present study inferred the proxy-response bias of total trip rates, the model could be applied to investigations of the bias in considering trip purpose, travel mode, and trip distance. Commuting and school trips should not be misreported by proxy-response but private and recreational trips and short-distance trips may be misreported by proxy-response. In addition, rounding errors for the trip departure or arrival time is another common bias observed in paper-based travel surveys. Such errors can be worsened by proxy response, and investigating these problems with our model should prove interesting.

Our present model assumes that only one household member will complete for all other members in the household. This assumption neglects the situation where all members in a household self-respond, or two or more members self-respond. We need to extend our model to relax this assumption. In the extended model, the task allocation model in existing group-decision modeling for household-based travel behavior analysis will be useful (e.g. Srinivasan and Bhat 2005, 2006; Weiss and Habib 2018; Zhang et al. 2009). Having access to actual completer data enables us to relax the one-completer assumption and expand the model structure such as Nested Logit or Mixed Logit.

The proxy response in this research is based on inference. Most Japanese household travel surveys do not include the completer information, and we have no available data for investigating the actual proxy response bias in Japanese household travel surveys. This is the one of the motivations of the present study. However, real proxy-response data is essential for examining the validity of our model. We will investigate survey data using completer information in future studies. In addition, our present model introduced only the basic information of individual characteristics, namely age and gender. If the completer data is available and we extend the model formulation, we can include other individual characteristics, such as occupation, health status, and education. For example, in an individual-based binary-logit model of proxy response, Ellis et al. (2003) use the independent variables including age, race, marital status, education attained, gender, activities of daily living, and health conditions, among others. This extension will be among our future work.

Besides these limitations, our proposed method will be useful for analyzing and correcting proxy-response bias, even if the completer is unknown, which is often the case when analyzing previously collected data. Preferably, the travel survey method should include the true completer information. However, existing travel survey data often lack the true completer information. Therefore, our proposed method is highly useful for inferring the proxy respondent and proxy-response bias.

As a final goal of our research project, we want to develop a bias-correction method of proxy response using our developed method. A promising method for performing bias correction is the propensity score analysis method implemented extensively in medical studies (e.g. Elliott et al. 2008; Ellis et al. 2003; Li et al. 2015). However, most existing studies of propensity score analysis use the individual-based choice model to calculate the probability of proxy response. Our proposed method using the group-based model will refine the probability and is expected to provide a sounder comparison of proxy and self-response and bias correction.

Acknowledgements This work was supported by JSPS KAKENHI, Grant Number JP18H01561. We would like to express our gratitude for this support. The authors thank Junyi Zhang, Eiji Hato, Morito Tsutsumi, and Makoto Chikaraishi for their insightful comments on an earlier version of this paper. The authors are also grateful for feedback from the participants at our presentations at hEART 2018, HKSTS 2018, and ARSC 2018. The comments by the anonymous reviewers were very helpful to improve the presentation of the paper. The remaining errors are the authors' responsibilities.

Authors' contributions TM: Conceptualization, literature review, manuscript writing, and editing. KH: Data preparation and model estimation. TK: Manuscript writing and data preparation.

Compliance with ethical standards

Conflicts of interest On behalf of all authors, the corresponding author states that there are no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A1

Our proposed method and the household utility model by Zhang et al. (2009) are equivalent. The choice probability in the maximum-type household utility function is given by

$$P_{i,0} = \frac{\exp V_{i,0}}{\exp V_{i,0} + \exp\{\max(V_{i,j} | j = 1, \dots, J_i)\}}, P_i^1 = 1 - P_{i,0} = \frac{\exp\{\max(V_{i,j} | j = 1, \dots, J_i)\}}{\exp V_{i,0} + \exp\{\max(V_{i,j} | j = 1, \dots, J_i)\}}, \quad (9)$$

where $P_{i,0}$ and P_i^1 are household-based non-response and response probabilities, respectively. If we approximate the maximum-type household utility function using the logsum variable, we have

$$P_{i,0} = \frac{\exp V_{i,0}}{\exp V_{i,0} + \exp(\ln \sum_{j=1}^{J_i} \exp V_{i,j})}, P_i^1 = 1 - P_{i,0}. \quad (10)$$

Zhang et al. (2009) call this expression the maximum utility (MAX) model. It is straightforward to show

$$P_{i,0} = \frac{\exp V_{i,0}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}}, P_i^1 = \sum_{j=1}^{J_i} P_j, P_{i,j} = \frac{\exp V_{i,j}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}} \text{ for } j \neq 0. \quad (11)$$

Thus, the MAX model by Zhang et al. (2009) and our proposed model are equivalent.

It should be noted that we usually require a choice indicator to identify the survey completer in the household if we estimate the model in Eq. (11) (or Eq. (1)) as the standard multinomial logit. However, if we estimate the binary choice model in Eq. (10), which is equivalent to the multinomial model in Eq. (11), we only require the binary choice indicator of household-based responses or non-responses. This is how the maximum likelihood in Eqs. (2)–(3) can estimate the parameters in Eq. (1) using the binary choice indicator δ_i .

Appendix A2

Here, we describe the details of the model by Maruyama (2014). This model assumes that no proxy response exists and every member in a household responds by themselves when a household sample is collected. Maruyama (2014) estimated an individual-based binary logit model (Table 5). Compared with this model, the group-based model (Table 1) exhibited an increase in the statistically significant variables and goodness-of-fit.

Table 5 Individual-based binary logit estimation of survey participation. *Source* Maruyama (2014)

Variables	Estimates	t-stat.
<i>Individual and household characteristics</i>		
Males under 39	−0.329	−9.56***
Females under 39	−0.239	−6.97***
Male 50+	0.285	6.53***
Female 50+	0.258	6.12***
Single household	−0.729	−15.81***
<i>Land use regulation in residential zone</i>		
Low-rise exclusive residential districts	0.403	3.18***
Medium-rise exclusive residential districts	0.123	2.49***
Category 2 residential districts	0.102	1.08
Quasi-residential districts	−0.694	−1.40
(Neighborhood) commercial districts	0.053	0.76
(Quasi-) industrial districts	−0.085	−0.95
No land use regulation	−0.168	−3.61***
Eastern-Ward	0.054	1.87
Constant (survey participate)	−0.686	−17.65***
Observations (individuals)		37,232
ρ^2		0.127
Adjusted ρ^2		0.126

***1% significant

References

- Ashley, D., Richardson, T., Young, D.: Recent information on the under-reporting of trips in household travel surveys. In: 32nd Australasian Transport Research Forum (2009)
- Badoe, D.A., Steuart, G.N.: Impact of interviewing by proxy in travel survey conducted by telephone. *J. Adv. Transp.* **36**, 43–62 (2002). <https://doi.org/10.1002/atr.5670360104>
- Bhat, C.R., Pendyala, R.M.: Modeling intra-household interactions and group decision-making. *Transportation* **32**, 443–448 (2005). <https://doi.org/10.1007/s11116-005-6789-x>
- Bhat, C.R., Goulias, K.G., Pendyala, R.M., Paleti, R., Sidharthan, R., Schmitt, L., Hu, H.H.: A household-level activity pattern generation model with an application for Southern California. *Transportation* **40**, 1063–1086 (2013). <https://doi.org/10.1007/s11116-013-9452-y>
- Bose, J., Giesbrecht, L.: Patterns of proxy usage in the 2001 national household travel survey. In: JSM Proceedings, Survey Research Methods Section, American Statistical Association. pp. 3313–3319 (2004)
- Chung, B., Srikukenthiran, S., Nurul, K., Miller, E.J.: The development of a web-survey builder (STAISI): designing household travel surveys for data accuracy and reduced response burden. In: 11th International Conference on Transport Survey Methods (2017)
- Cobb, C.: Answering for someone else: proxy reports in survey research. In: Vannette, D.L., Krosnick, J.A. (eds.) *The Palgrave Handbook of Survey Research*, pp. 87–93. Palgrave Macmillan, Basingstoke (2018a)
- Cobb, C.: Proxy reporting. In: Vannette, D.L., Krosnick, J.A. (eds.) *The Palgrave Handbook of Survey Research*, pp. 427–437. Palgrave Macmillan, Basingstoke (2018b)
- de Palma, A., Picard, N., Inoa, I.: Discrete choice decision-making with multiple decision-makers within the household. In: Hess, S., Daly, A. (eds.) *Handbook of Choice Modelling*, pp. 363–382. Edward Elgar, Cheltenham (2014)
- Elliott, M.N., Beckett, M.K., Chong, K., Hambarsoomians, K., Hays, R.D.: How do proxy responses and proxy-assisted responses differ from what medicare beneficiaries might have reported about their health care? *Health Serv. Res.* **43**, 833–848 (2008). <https://doi.org/10.1111/j.1475-6773.2007.00820.x>
- Ellis, B.H., Bannister, W.M., Cox, J.K., Fowler, B.M., Shannon, E.D., Drachman, D., Adams, R.W., Giordano, L.A.: Utilization of the propensity score method: an exploratory comparison of proxy-completed to self-completed responses in the Medicare Health Outcomes Survey. *Health Qual. Life Outcomes*. **1**, 47 (2003). <https://doi.org/10.1186/1477-7525-1-47>
- Ermagun, A., Levinson, D.: Intra-household bargaining for school trip accompaniment of children: a group decision approach. *Transp. Res. Part A Policy Pract.* **94**, 222–234 (2016). <https://doi.org/10.1016/j.tra.2016.09.012>
- Gliebe, J.P., Koppelman, F.S.: Modeling household activity-travel interactions as parallel constrained choices. *Transportation* **32**, 449–471 (2005). <https://doi.org/10.1007/s11116-005-5328-0>
- Ho, C., Mulley, C.: Intra-household interactions in transport research: a review. *Transp. Rev.* **35**, 33–55 (2015). <https://doi.org/10.1080/01441647.2014.993745>
- Ho, C.Q., Mulley, C., Shiftan, Y., Hensher, D.A.: Vehicle value of travel time savings: evidence from a group-based modelling approach. *Transp. Res. Part A Policy Pract.* **88**, 134–150 (2016). <https://doi.org/10.1016/j.tra.2016.03.016>
- Li, M., Harris, I., Lu, Z.K.: Differences in proxy-reported and patient-reported outcomes: assessing health and functional status among medicare beneficiaries. *BMC Med. Res. Methodol.* **15**, 62 (2015). <https://doi.org/10.1186/s12874-015-0053-7>
- Maruyama, T.: Participation choice model for household travel survey methods: Comparison of paper, web, and smartphone-based methods. In: Proceedings of the 19th International Conference of Hong Kong Society for Transportation Studies (2014)
- Moore, J.C.: Self/proxy response status and survey response quality, a review of the literature. *J. Off. Stat.* **4**, 155–172 (1988)
- Reynolds, J., Wenger, J.B.: He said, she said: the gender wage gap according to self and proxy reports in the current population survey. *Soc. Sci. Res.* **41**, 392–411 (2012). <https://doi.org/10.1016/j.ssresearch.2011.10.005>
- Richardson, A.J.: Proxy responses in self-completion travel diary surveys. *Transp. Res. Rec. J. Transp. Res. Board.* **1972**, 1–8 (2006). <https://doi.org/10.1177/0361198106197200101>
- Seebauer, S., Fleiß, J., Schweighart, M.: A household is not a person: consistency of pro-environmental behavior in adult couples and the accuracy of proxy-reports. *Environ. Behav.* **49**, 603–637 (2017). <https://doi.org/10.1177/0013916516663796>
- Skolarus, L.E., Sánchez, B.N., Morgenstern, L.B., Garcia, N.M., Smith, M.A., Brown, D.L., Lisabeth, L.D.: Validity of proxies and correction for proxy use when evaluating social determinants of health in stroke patients. *Stroke* **41**, 510–515 (2010). <https://doi.org/10.1161/strokeaha.109.571703>

- Srinivasan, K.K., Athuru, S.R.: Analysis of within-household effects and between-household differences in maintenance activity allocation. *Transportation* **32**, 495–521 (2005). <https://doi.org/10.1007/s11116-005-5355-x>
- Srinivasan, S., Bhat, C.R.: Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation* **32**, 523–544 (2005). <https://doi.org/10.1007/s11116-005-5329-z>
- Srinivasan, S., Bhat, C.R.: A multiple discrete-continuous model for independent- and joint-discretionary-activity participation decisions. *Transportation* **33**, 497–515 (2006). <https://doi.org/10.1007/s11116-006-8078-8>
- Stopher, P.: *Collecting, Managing, and Assessing Data Using Sample Surveys*. Cambridge University Press (2012)
- Timmermans, H.J.P., Zhang, J.: Modeling household activity travel behavior: examples of state of the art modeling approaches and research agenda. *Transp. Res. Part B Methodol.* **43**, 187–190 (2009). <https://doi.org/10.1016/j.trb.2008.06.004>
- Todorov, A.: Cognitive procedures for correcting proxy-response biases in surveys. *Appl. Cogn. Psychol.* **17**, 215–224 (2003). <https://doi.org/10.1002/acp.850>
- Todorov, A., Kirchner, C.: Bias in proxies' reports of disability: data from the National Health Interview Survey on disability. *Am. J. Public Health* **90**, 1248–1253 (2000). <https://doi.org/10.2105/ajph.90.8.1248>
- Verreault, H., Morency, C.: What about proxy respondent bias over Time? CIRRELT-2015-55 (2015)
- Verreault, H., Morency, C.: Integration of a phone-based household travel survey and a web-based student travel survey. *Transportation* **45**, 89–103 (2018). <https://doi.org/10.1007/s11116-016-9726-2>
- Wagmiller, R.L.: A fixed effects approach to assessing bias in proxy reports. *Int. J. Public Opin. Res.* **21**, 477–505 (2009). <https://doi.org/10.1093/ijpor/edp035>
- Wargelin, L., Kostyniuk, L.: Proxy respondents in household travel surveys. In: Stopher, P., Stecher, C. (eds.) *Travel Survey Methods*, pp. 201–212. Emerald Group Publishing Limited, Bingley (2006)
- Weiss, A., Habib, K.N.: A generalized parallel constrained choice model for intra-household escort decision of high school students. *Transp. Res. Part B Methodol.* **114**, 26–38 (2018). <https://doi.org/10.1016/j.trb.2018.05.008>
- Zhang, J., Kuwano, M., Lee, B., Fujiwara, A.: Modeling household discrete choice behavior incorporating heterogeneous group decision-making mechanisms. *Transp. Res. Part B Methodol.* **43**, 230–250 (2009). <https://doi.org/10.1016/j.trb.2008.05.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Takuya Maruyama is currently an associate professor at Kumamoto University, Japan. He received his Bachelor's, Master's, and Doctoral degrees from the University of Tokyo, Japan. He specializes in travel demand forecasting, network equilibrium models, and travel survey methods.

Kenta Hosotani graduated from Kumamoto University with his Bachelor's degree in civil engineering in 2018. He has been working at Kumamoto Prefectural Government.

Tomoki Kawano graduated from Kumamoto University with his Master's degree in civil engineering in 2019. He has been working at Nippon Koei Co., Ltd.