The mixing of survey modes: application to Lyon web and face-to-face household travel survey

Caroline BAYART, Laboratoire d'Economie des Transports
Patrick BONNEL, Laboratoire d'Economie des Transports

12th World Conference on Transportation Research
Lisbon, 11-15 July 2010

Authors

Caroline BAYART
Laboratoire d'Economie des Transports
Université Lyon 2 - ENTPE
France
Phone: 00(33)4.72.04.72.48
Fax: 00(33)4.72.04.70.92
E-mail: caroline.bayart@entpe.fr

Patrick BONNEL
Laboratoire d'Economie des Transports
Université Lyon 2 - ENTPE
France
Phone: 00(33)4.72.04.70.48
Fax: 00(33)4.72.04.70.92
E-mail: patrick.bonnel@entpe.fr

The mixing of survey modes: application to Lyon web and face-to-face household travel survey

Caroline BAYART, Laboratoire d'Economie des Transports
Patrick BONNEL, Laboratoire d'Economie des Transports

Abstract

As response rates with all traditional modes (face-to-face, telephone and postal) are declining, it becomes difficult to carry out efficient households travel surveys. Non respondents probably have different behaviour from those who agree to be interviewed. To reduce this bias of non-response, we initiated a project of a web survey in parallel of the household travel survey conducted in face to face in Lyon in 2006. The idea is to propose to those households who refuse to respond or are not contactable after a certain number of attempts to respond by the web. The two main objectives of this research are to test the feasibility of a web survey for non-respondents and compare mobility results of both survey modes. We focus on the main bias often meet with this new media: on-line respondents belong to specific categories. After a description of the population who answered on-line, we characterize its travel pattern and estimate a selection bias.

Keywords

Household travel surveys, web surveys, mixed modes surveys, data comparability, response rate, sampling coverage, design issues, selection bias.

Introduction

Survey response rates are decreasing over the world (Atrostic, Burt, 1999). Household travel surveys follow the same tendency. Even if weighting procedure allows to reduce the incidence of non-response, it is always necessary to postulate that people with some socio-demographic characteristics who do not respond to a survey have the same behaviour than people with the same socio-demographic characteristics who respond. But evidence seems to indicate that it is not the case for travel (Richardon, Ampt, 1993; Richardson, 2000, Ampt 1997). Therefore, survey non-response might produce bias. Some efforts are put to increase response rate for traditional survey by improving the questionnaire, reducing respondent burden, increasing reminders… Even if results are generally positive, it is in most cases not sufficient. According to Alsnih (2004), a way to increase the response rate and produce more reliable results is to propose a second media to answer. That's why we initiated a project of a web survey in parallel of the household travel survey conducted in Lyon. The main survey is still realised with the same methodology as for previous survey (CERTU, 1998), households being interviewed at home in face to face (11.000 interviews have been done between October

2005 and March 2006). The idea is to propose to households who refuse to respond (about 18% in 1995) or are not contactable after a certain number of attempts (about 17% in 1995) to respond by the web. Those households are informed by a letter of this option, followed by two reminders.

This new and interactive mode of data collection offers to the respondents the possibility to choose a nice moment to complete the questionnaire, and does not require setting an appointment with the interviewer. However, Internet penetration rate is still low, and users' capabilities and equipment vary a lot. If Web surveys allow to reduce the non-response rate, the generalization of the results to the whole population remains inaccurate (Myles & Tibert, 1998). Moreover, the implementation of a Web survey raises specific problems, in terms of design and administration of the questionnaire. Lyon households travel survey is traditionally administered in face-to-face by experimented interviewers. Propose an on-line survey requires an auto-administrated questionnaire. That's why it's important to work on its attractiveness, its simplification and its technical feasibility to not discourage potential respondents. The task is particularly complex with regard to the gathering of individual trips. Lastly, if the launch of a web survey makes it possible to study behaviours little represented up to now (hyper-mobiles households, with shifted schedules...), the question of data comparability remains entire (Stopher & Jones, 2003). The danger when databases are merged is that a sample selection bias will be created that will compromise the accuracy of explanatory models of travel behaviours.

This paper initially explains the Lyon web travel survey methodology (section 1) and gives some thoughts on respondents' profile and mobility pattern. Comparisons between web and face-to-face surveys are provided (section 2). Then, we characterize a selection bias (section 3). Finally, we give some perspectives for future households travel surveys (section 4).

1. Methodology

The methodology for French urban household travel surveys has been laid down by the CERTU (Certu 2008). Traditionally, such surveys have been conducted face-to-face. The most recent survey in Lyon, which dates from 2006, involved a random geographically stratified sample drawn from the France Télécom telephone directory. Those conducting the survey identified other addresses in the field, starting from the initial list of addresses, by following a precise rule (route sampling). This method results in a random sample that is representative of the population of the study area, overcoming problems to do with the representativeness of the households in the telephone directory. Those individuals who refused to respond to the standard survey, or who were unreachable or absent for long periods, were asked to answer the on-line questionnaire by letter as we did not have their e-mail address and did not always have their telephone number (Figure 1). The perimeter of the web-based survey took in the 72 municipalities covered by the Lyon Schéma de COhérence Territorial (S.C.O.T.) which is the master plan for the conurbation.

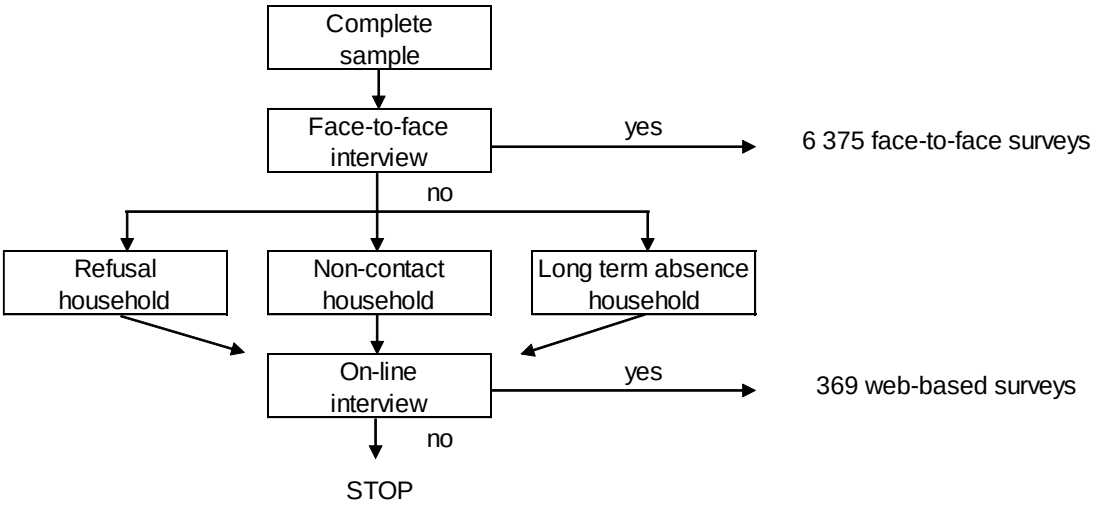


Figure 1: Household recruitment

Face-to-face interviews took place between November 2005 and May 2006. At the end of the period, 6,375 interviews were realized on the survey perimeter. As 11,951 households have been contacted, the response rate is around 53 %. Households who didn't want to welcome an interviewer at home received a letter which asked them to fill out the questionnaire by Internet. A similar letter was also sent to the households not contactable by phone after 8 attempts. These letters especially insisted on the new media available to participate to the survey, which is more flexible. They were sent by mail because the electronic address of the selected households was unknown. It mentioned the accurate address of the website where the questionnaire is available, a moderate response time (20 minutes), as well as a personal login and password. Two waves of sending have been carried out (1,882 at the beginning of April and 2,497 in the middle of May 2006), in order to limit the time between the refusal and the revival. Then, the survey administrator sent two reminders per household, between mid-May and the end of June 2006.

The questionnaire had a three parts structure for both survey modes (web-based and face-to-face). It began with the questions about the household, which were followed by those about the respondent, and ended with those about all the trips made on the day before the survey was conducted. Only weekdays were covered in the survey. Respondents do not always find it easy to understand or absorb the concept of trip, which is central to the study. To guarantee data quality and avoid measurement biases due to different interpretations of this concept during the face-to-face interviews, the interviewers explained it to the respondents using their instruction manual. This is impossible during a web-based survey as it is self-administered. There is no certainty that when they responded to the questionnaire all the Internet users understood and applied the definition of trip that was provided on line, i.e. that given by the CERTU in the methodological guide (Certu 2008). It was therefore necessary to guide Internet respondents to help them transcribe their trip sequence. In concrete terms, the Internet respondents began by precisely describing where they were at the start of the travel day (when the addresses of their home and place of work or study had already been entered these were generated automatically), and what activity they performed there. The questions were then asked in the order shown in Figure 2, until the departure time for the last trip was later than the end of the survey period. We have continued to use trips rather than activities as an input in spite of the findings of a considerable amount of research (Stopher 1998) to obtain better comparability with the face-to-face interview.

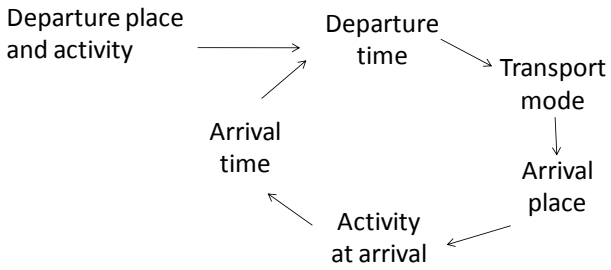


Figure 2: Daily Trips Pattern
Source: Web questionnaire developed for the Lyon household travel survey (2005)

How respondents perceive the questions and how the questionnaire is organized depend on the capabilities of the survey medium. The questionnaire used for the face-to-face interview was therefore simplified for the self-administered web-based survey. In addition, a single member over 11 years old of the household filled in the on-line questionnaire, while all the family members were questioned during the face-to-face interview (Bayart & Bonnel, 2008).

2. Results of the fieldwork

The implementation of an on-line questionnaire is worth not only for the survey administrator, but also for all the households who are familiar with Internet. But, on-line respondents are belonging to specific categories (young person, male, with a high level of education and incomes, large household …). Moreover, if we suppose that socio-economic characteristics of households who answer on Internet differ from those of the target population, we can think that their mobility habits are not the same (Resource System Group, 2002). First, we present the general results of the survey, in terms of response rate and response behaviour. Then, we describe the population who answered on-line, and characterize its travel pattern.

2.1. Response rate and response behaviour

As of July 10th, 2006, around 536 individuals reached the site to try to answer the survey. But only 369, which represents an answer rate of 8.5 %, gave a complete and exploitable answer. The follow-up appears successful, because it generates around 140 completed on-line interviews. The results are set out in Table 1.

Table 1: Response rate according to survey mode

Survey mode	Face-to-face	Web
Households contacted	11,951	4,335
Households interviewed	6,375 (53%)	536 (12.5%)
Completed interviews	6,375 (53%)	369 (8.5%)

These results seem small, compared to face-to-face survey, but are promising if we consider that households have refused to participate to the face-to-face interview and that some of them are not able to connect to Internet at home or at their working place (50% could be considered as net surfers according to recent official statistics).
The average time required to answer the Web questionnaire is close to the one announced in the contact letter. Disparities are however important (between 5 mn and more than one hour and a half). The longest part concerns the report of daily trips (more than 9 mn on average).

This is due to the number and the precision of the questions, which ask for an important memory effort (addresses of visited locations, departure and arrival times…). For the other parts of the questionnaire, response times are shorter (between 2.5 and 4 mn on average). "Household" and "individual" blocks include more "usual" questions for interviewees, and often suggest drop-down lists for the answers. This is simple for the Internet respondent, who minimizes his connection time. During the week, net surfers connect from 9:00 am till 11:00 pm, with three different peaks: between 11:00 am and 1:00 pm, between 3:00 pm and 5:00 pm and between 7:00 pm and 10:00 pm. It is likely that numbers of connections are realized in the workplace, as far as 20 % of Web respondents declare not to have internet access at home. The weekend, one connection peak appears after 6:00 pm.

2.2. Who are web respondents?

Data comparison of standard household's survey and Web survey allows us to raise the profile of web respondents. If differences in terms of residential location are not significant, other socio-economic characteristics stand out clearly with statistically significant differences at 5 % level.

The households having answered the Web survey have more members (2.6 persons vs. 2.4). We can suppose that it is due to families, although we do not have all the data on the composition of the household. The proportion of 4 persons' households is indeed more important in the Web survey (22 %, vs. 13 %), while the proportion of single person's households is weaker (24 %, vs. 32 %). Internet households are more motorized. Half of them possess at least two cars, against only a third in the standard survey. If we look at the number of vehicles by person in age to drive, conclusion is the same. This larger car access is partially explained by the size of the household and the socio-economic characteristics of the Web respondents. They are also better equipped in communication tools, because 80 % have an internet access at home and a mobile phone (vs. respectively 51 % and 69 % in the standard survey).

They have a higher annual income by unit of consumption (20,000 euros vs. 15,000 euros on average). This difference can be explained by the fact that the penetration of internet is still uneven on the territory, and concerns first high incomes households (cost of the investment in computer hardware, internet subscription, and higher level of study…). Moreover, those who have a connection at the workplace have generally superior occupations, with a higher level of income. Finally, the part of households refusing to give their annual income level is less important on Web (1 household on 4, vs. 1 household on 3). Respondents are reluctant to communicate personal information, especially when it deals with their incomes. It seems more difficult to declare its incomes in face-to-face to an interviewer than on the Web, which is by definition auto-administered.

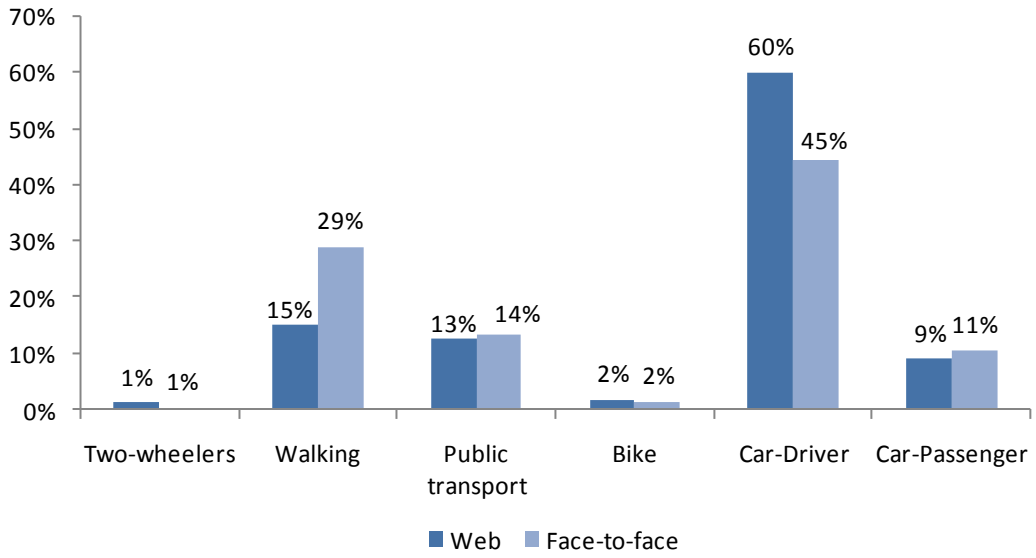
Individual data analysis allows us to refine this comparison. The part of women having answered on Web is slightly more important than the one observed in face-to-face, but the difference is not significant. On the other hand, the proportion of active persons (between 30 and 59 years) having answered on Web is larger (three quarters, vs. 50 % in the standard survey). They belong to high social and occupational groups (39 % are executives, or have a superior intellectual profession, vs. 14 %) and have a high level of studies (about 2/3 made studies vs. only 30 % in the standard survey). These two variables are linked, and are explained partially by the person who chose to fill out the questionnaire (the household leader or his spouse). Web respondents have more frequently a driving license (93 %, vs. 81 %). Oldest and youngest persons are under-represented in the Web survey. Finally, they work or study more often in the town centre (27 %, vs. 16 %), where the head offices of companies are located. This population has particular characteristics. Active persons may be less available to

answer a long questionnaire in face-to-face. Probably more sensitive to the survey subject and familiar with internet, they agreed to answer the on-line questionnaire. These households are nevertheless more difficult to contact by phone, because less than 60 % declare to be registered on the list of the France Telecom subscribers.

2.3. Mobility pattern

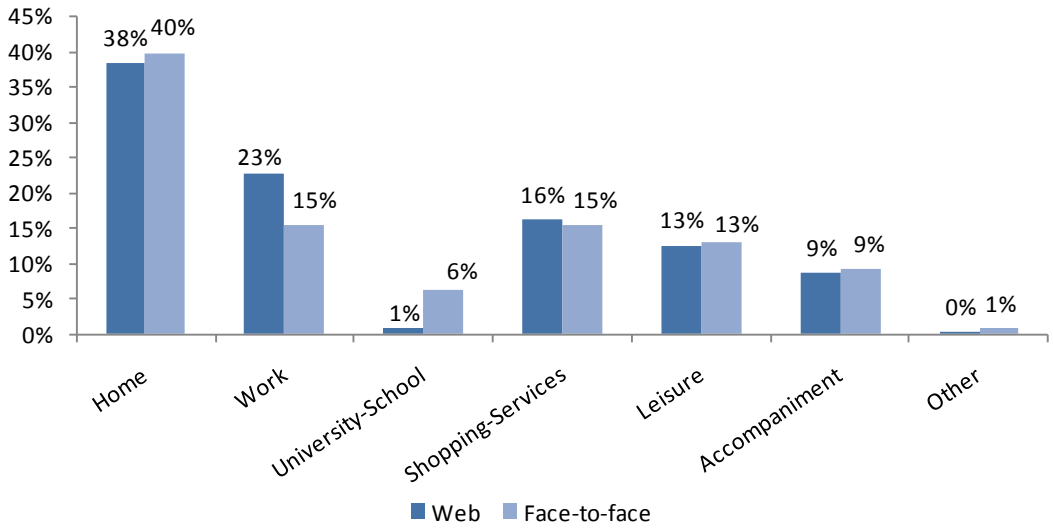
Having qualified both samples on the basis of socio-economic characteristics, we are going to analyze the mobility pattern of these respondents. The objective of this part is to show the differences in terms of behaviour between Web and face to face samples. It is indeed strongly likely that socio-economic differences have an impact on mobility pattern.

Web respondents declare fewer daily trips than the one interviewed in face-to-face (3.6 vs. 3.0). This gap is explained by a more important number of immobiles (19 % vs. 11 %) and a weaker mobility of mobiles in the Web survey (3.7 vs. 4.1 daily trips). The administration mode of the questionnaire also impacts these results. At home, an interviewer can send reminders and verify the relevance of the trips pattern. On Web, the respondent is alone in front of his computer, and can easily omit "small" trips.



Source: Households web and face-to-face travel survey - Lyon 2006
Figure 3: Distribution of daily trips by transportation mode and by survey

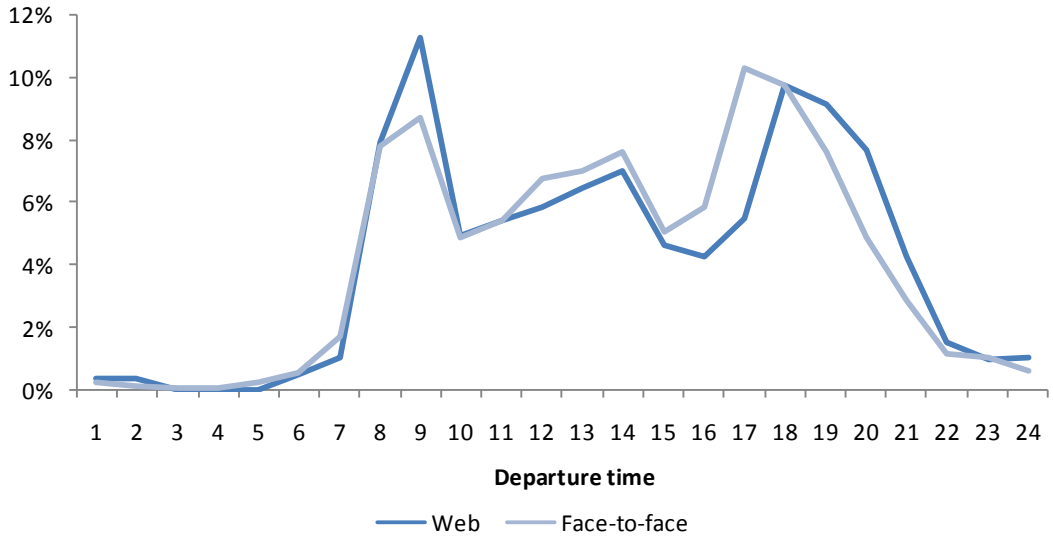
They use more often a car (58 % vs. 44 %), and less walking (15 % vs. 29 %) (Figure 3). But if we add the data concerning all motorized modes, the mobility of Web respondents becomes superior to the one of face-to-face respondents, especially if we exclude those who did not move (2.6 vs. 2.3 trips per day). It would thus seems that Web survey allows getting back "correctly" motorized trips of mobile persons, but leads to a strong underestimate of walking trips, which are also “small” trips. The purpose at destination "Work" is over represented in Web survey, with regard to the standard survey (23 % vs. 15 %), because three quarter of the web respondents are active persons (Figure 4).



Source: Households web and face-to-face travel survey - Lyon 2006
Figure 4: Distribution of daily trips by purpose at the destination and by survey

Web respondents cover on average longer distances (4.6 km vs. 3.8 km), and their trips have higher duration (25.5 mn vs. 19.3 mn). In Web survey, we noticed a strong proportion of trips made in a private car from home to work, but few walking trips.

That's why declared distances and lengths of trips in the area of study are higher on average. Finally, web respondents seem to come back to home later in the evening (Figure 5).



Source: Households web and face-to-face travel survey - Lyon 2006
Figure 5: Average number of daily trips by departure time

Lastly the number of daily tours from home is slightly weaker for Web respondents (1.2 vs. 1.5). 68 % make only a single tour, vs. 61 % in the face-to-face survey. Average tours' distances and lengths are higher on the web survey (11 vs. 9 km and 64 vs. 50 mn).

These results can be explained by two main factors: socio-economic differences (share of active persons, driving license, car ownership, level of income and studies, age, status) and an

effect of the survey mode, which can lead to an under report of the number of daily trips. In the following subsection, we are going to partially neutralize the socio-economic characteristics of the respondents. The issues which then arise are the comparability of the data and the methodologies which can be used to separate those differences in responses which are due to the methodology from real differences in the behaviour of each group.

3. Sample selection bias

It has been known since the 1950s that estimating an equation on the basis of a subsample selected from the population may result in biases (Roy, 1951). However, the first econometric exploration of the consequences of such sample selection was Heckman's work in 1974. The standard example is that of the estimation of salary based on an analysis of working women on their own, because the decision to work involves a trade-off in which the salary an individual can potentially earn plays a role. Since this, many papers have highlighted the importance of the selection bias in human and social science surveys (Maddala, 1986). Noteworthy examples are the model for migration in the USA analyzed by Nakosteen and Zimmer (1980), or that for female employment rates analyzed by Mroz (1987). The most frequent use of self-selection models is for the evaluation of processing or training.
In practice, the selection bias has two sources (Heckman, 1990). It results either from respondent self-selection or a selection decision by the study managers. When mixed survey modes are used, individuals choose to belong to one group or another or only respond if the proposed medium suits them. The responses are therefore not comparable, because the sample is no longer random and the presence of respondents is determined by external factors which may also affect the variable of interest in the studied model. It is highly likely that the socioeconomic characteristics and the travel behaviours of the individuals who respond using the Internet are different from those of the individuals who respond to a face-to-face interview (Resource system group, 2002; Lozar Manfreda and Vehovar, 2002).

The data from the Lyon household travel survey highlight the problem of self-selection, with the non respondents to the standard face-to-face survey choosing whether or not to fill in the Internet questionnaire. We shall illustrate this by proposing an explanatory model for travel. More precisely, we shall analyze the average number of trips per person (Bayart et al., 2009). Our analysis shows that the Internet respondents travelled less than the face-to-face respondents (3.0 vs 3.6 daily trips). Let us consider an equation that permits an analysis of the effect of survey mode on an individual's average number of daily trips:

$$Y_i = \beta_k X_{ki} + \alpha I_i + u_i$$

where Y_i is the average number of trips made by respondent i (dependent variable), X_{ki} is a vector of explanatory variables and I_i is a dummy variable that states whether the individual I responded by Internet. A question which arises is whether the coefficient α measures the real impact on daily travel of the choice of responding on the Internet. The answer to this question is affirmative if individuals who decide to respond on the Internet would have reported the same number of trips if they had responded in the face-to-face situation. However, the variable I cannot be considered as exogenous in this model, as the contacted individuals chose whether or not to respond in face-to-face interview and whether or not fill out the on-line survey. Respondent self-selection must be corrected during least squares regression in order to obtain unbiased estimates of the coefficients, by using the two-stage estimation method developed by Heckman in 1979.

3.1. Selection equation

The first stage consists of estimating the survey medium "choice" equation using a probit model. Before commencing the econometric analysis, we shall select variables which make it possible to distinguish the individuals who responded on the Internet. Two types of variables drew our attention in this context: the household's characteristics as regards telecommunications (possession of an Internet connection at home or a cell phone and presence in the directory) and some sociodemographic characteristics (status, age, educational level, whether they stated their annual income, the number of persons in household and place of work).

The coefficients estimated by the probit model are significant (table 2). An examination of their sign gives a good idea of the importance of the various factors which influence choice of the Internet. Unsurprisingly, the possession of an Internet connection in the home or a cell phone, and the absence of a subscriber from the directory increase the probability of responding to the survey by Internet. Likewise, the probability of responding on line increases when individuals report their annual income, work in the city centre and have gone through higher education. Conversely, the probability of using the Web falls with household size and for individuals currently in education.

Table 2: Selection model (Probit)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.408	0.170	-14.15	< 2e-16	***
Age	0.0046	0.0022	2.09	0.036	*
Internet connection at home	0.424	0.069	6.17	7.0e-10	***
Ownership of mobile phone	0.331	0.074	4.43	9.4e-06	***
Telephone_no directory	0.370	0.056	6.62	3.6e-11	***
Telephone_no fixed line	0.317	0.094	3.37	0.00074	***
Occupation_working unstated location	-0.859	0.269	-3.19	0.0014	**
Occupation_working in suburbs	-0.187	0.063	-2.96	0.0031	**
Occupation_not working	-0.291	0.077	-3.77	0.00016	***
Number of individuals in household	-0.098	0.020	-4.84	1.3e-06	***
Declared income	0.294	0.060	4.91	9.0e-07	***
Educational level_ongoing	-0.387	0.129	-3.01	0.0026	**
Educational level_not higher	-0.358	0.058	-6.21	5.2e-10	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Sample size: 13 483

3.2. Regression model

The second stage consists of explaining the differences in travel behaviour using a specific model which includes:

• A dependent variable, the average number of trips made by individuals;
• Several independent or explanatory variables, the observed factors assumed to have an effect on the number of trips reported by individuals;
• The inverse Mills' ratio, a correction factor for each individual obtained in the first stage;

• An error term which takes account of the unobserved forces that could influence the measurement of results.

What we will do is estimate two models, one using the subsample of face-to-face respondents, and the other using the subsample of internet respondents. The coefficients of the explanatory variables and the selection bias variable are estimated by a least squares regression method. We shall only consider here the explanatory variables with a direct impact on individuals' travel (gender, age group, residential location, the number of individuals and children in the household, possession of a driving license, the number of cars owned by the household, the day the individuals responded to the questionnaire and willingness to report their annual income). In each model, the coefficient of the selection bias variable is significant. Individuals are therefore subject to a selection bias, and this can be eliminated by using a two-stage estimation method. We thus obtain unbiased estimates.

Table 3: Travel model for face-to-face respondents (OLS)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.56e+00	2.33e-01	10.97	< 2e-16	***
Gender_male	-1.44e-01	4.17e-02	-3.45	0.00055	***
Possession of driving licence	6.65e-01	6.22e-02	10.69	< 2e-16	***
No. cars per person ≥ 18 years	5.15e-01	6.00e-02	8.58	< 2e-16	***
Reference day Friday	1.93e-01	5.70e-02	3.39	0.00071	***
No. of children in household	7.40e-01	5.35e-02	13.84	< 2e-16	***
(No. of children in household)²	-5.53e-02	1.25e-02	-4.44	9.2e-06	***
No. of persons in household	-2.18e-01	2.48e-02	-8.78	< 2e-16	***
Residential location / centre	1.07e-05	4.20e-06	2.55	0.0108	*
Declared income	2.34e-01	4.65e-02	5.03	5.1e-07	***
Age	5.59e-02	5.67e-03	9.87	< 2e-16	***
$(Age)^2$	-6.97e-04	6.01e-05	-11.60	< 2e-16	***
Mills	-2.55e-01	6.25e-02	-4.07	4.6e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Sample size: 13 155

The model which is restricted to the subsample that responded to the face-to-face survey provided interesting results, as all the variables are significant (table 3). Being a man and belonging to a large household has a negative impact on the propensity to travel. In contrast, living far from the centre, having cars available and having a driving license increases the average daily number of trips. Travel also seems to be an increasing function of age up to 40 years of age (the square of age is also included in the model to take account of the decrease in travel among individuals over the age of 40 to 50) and of the number of children in the household. The Mills' coefficient has a negative sign: those who responded to the face-to-face survey would have declared a lower number of trips if they had answered the web survey.

In the model which is restricted to the subsample of individuals who responded on line, few of the coefficients achieved significance (table 4). This is explained in particular by the differences in the sizes of the two samples of respondents. A total of 13,271 individuals

responded to the face-to-face survey, compared with only 369 on the web, i.e. a ratio of 1 to 36. The order of magnitude of the standard deviation of the estimated coefficients therefore varies in a ratio of 1 to 6 between the Web and face-to-face samples. Being male, taking Friday as the reference for trips and living far from the centre have a negative impact on travel (generally women make more "shopping" and "accompanying" trips). However, the number of trips increases with household size. The Mills' coefficient has a negative sign: those who decided to answer on the Web would have reported more daily trips if they had answered to face-to-face interview.

Table 4: Travel model for Internet respondents (OLS)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.68e+00	1.60e+00	2.92	0.0037	**
Gender_male	-7.22e-01	2.52e-01	-2.86	0.0045	**
Possession of driving licence	8.83e-01	5.06e-01	1.74	0.0821	.
Cars per person ≥ 18 years	5.95e-01	3.61e-01	1.65	0.0999	.
Reference day Friday	-5.00e-01	2.47e-01	-2.02	0.0440	*
No. of children in household	7.80e-01	4.87e-01	1.60	0.1102	
(No. of children in household)²	-4.36e-01	1.74e-01	-2.51	0.0125	*
No. of persons in household	4.87e-01	1.59e-01	3.07	0.0023	**
Residential location / centre	-6.24e-05	2.72e-05	-2.29	0.0226	*
Declared income	2.68e-01	3.45e-01	0.78	0.4379	
Age	-4.09e-02	4.73e-02	-0.86	0.3878	
$(Age)^2$	3.72e-04	5.09e-04	0.73	0.4658	
Mills	-1.11e+00	4.16e-01	-2.66	0.0083	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Sample size: 328

3.3. Model Stability Test

The purpose of this is to test whether the estimated coefficients of the two stages model are stable and to quantify any impact of the survey mode. The results from the regression model applied to the explanatory variables and the significant interactions are set out in Table 5.
Three variables appear to interact significantly with the response mode. These are gender, household size and the distance between the household's residence and the city centre. However, the Mills ratio coefficients for the two samples do not differ significantly. In our analyses we shall retain those interactions that are significant at the 10 percent error level, with the exception of the coefficient of the mode variable (p-value = 17.6 percent).

Table 5 Stable Model with Interactions (OLS)

	Estimate	Pr(>|t|)	Sign.
(Intercept)	2.65e+00	<2e-16	***
Gender_male	-2.40e-01	2.08e-07	***
Possession of driving licence	4.453e-01	2.63e-09	***
Cars per person ≥ 18 years	5.34e-01	8.07e-15	***
No. of children in household	6.94e-01	<2e-16	***
(No. of children in household)²	-3.3e-02	0.051	.
No. of persons in household	-1.73e-01	1.93e-10	***
Residential location / centre	9.54e-06	0.042	*
Declared income	2.65e-01	2.84e-07	***
Age	5.7e-02	1.03e-11	***
$(Age)^2$	-6.36e-04	1.53e-12	***
Mills	-1.91e+00	6.15e-04	***
Mode	-5.04e-01	0.176	
(Gender_male) * mode	-5.47e-01	0.059	.
(No. of persons in household) * mode	1.87e-01	0.102	
(Residential location / centre) * mode	-5.9e-05	0.053	.

The data collection mode (web or face-to-face) had a direct impact on travel levels. The number of trips fell by 0.5 when subjects answered the questionnaire on line.
The coefficient of the variable "distance from the place of residence to the city centre x mode' was slightly negative (-5.9e-05). This means that the web respondents' number of trips decreased with the distance between their place residence and the city centre of. The impact of this was 9.5e-06 trips for respondents to the face-to-face survey vs. -4.9e-05 trips for respondents to the web-based survey. More of the internet respondents are employed as executives and office workers with jobs in the city centre. The further their home is from the city centre, the greater the distance between their home and their work. This reduces the time available for activities that are less constrained which reduces their levels of travel. It is striking that this coefficient is much lower than the others. The explanation for this lies in the unit used (metres) to calculate the distance between respondents' homes and the city centre.
Being male had a negative impact on the mobility of respondents to the web-based survey: its impact was -0.24 trips per day for face-to-face respondents vs. – 0.79 for internet respondents. As has been shown above, women make more trips than men as they have to manage family activities in addition to their work. This effect is amplified in the case of internet respondents. Perhaps this is because their more demanding jobs (a large number of them are executives) give them less scope for making trips. It should be borne in mind that three quarters of the internet respondents were economically active.
In contrast, the mobility of internet respondents increased with household size. The impact on the number of daily trips was -0.17 per person for face-to-face respondents, vs. 0.014 for internet respondents. The latter have higher annual incomes. It is likely that as the number of persons in a household increases the number of trips made by its members increases also (for shopping or leisure purposes, for example).
The coefficients of the variables (Gender:male)*mode, (No. of persons/household)*mode and (Distance between home and centre)*mode are thus able to quantify the impact of the survey mode on travel behaviour.

4. Conclusion

Transport constitutes a strategic tool for urban policies. Knowing with accuracy the travel pattern of inhabitants is necessary to ensure a durable development and management of infrastructures. The relevance of selected investments will depend on the quality of gathered data. It thus becomes crucial to identify a reliable survey protocol to obtain representative data at reasonable cost. In a perfect world, the methodology implemented would make it possible to survey the entire selected population according to a good quality/cost ratio, to reduce the non-response rate, to respect the cultural differences and to develop a reliable forecasting model (Morris et al., 2003). Unfortunately, it is not possible to reach all these objectives in a household travel survey. Recent technological advances, like the expansion of Internet, open new windows which have to be studied. To improve data quality, the Web is used for the first time in Lyon to support the traditional face-to-face method. The aim of the experiment was to capture some households who didn't want to respond to the traditional face-to-face survey. The web is not considered as an alternative media to fill out the questionnaire, but as a way to survey non respondents to traditional survey media. This method can increase survey quality by generating information usually ignored. It could be a way to balance the growth of refusals endured by interviewers since many years.
The gathering of data on-line also generates some comparability issues. The development of a new questionnaire and its differences compared to the one diffused in face-to-face (shorter, self-administrated, more filters…) may influence the answers (Dillman & Browker, 2001). One of our main objectives was to carry out a comparative analysis of mobility behaviours, between households who answer on-line and those subjected to the standard questionnaire. This analysis is rather complex, as far as we have to take into account of socio-economic differences between individuals who answered on the Web and those having accepted the face-to-face interview, differences of mobility pattern between these two sample and finally differences which can be imputed to the media only. Data show that Internet respondents move less than respondents to face-to-face survey. This slighter mobility concerns the daily number of trips and tours. This gap can be explained by a double effect: a higher immobility of the web respondents and a lower statement of their daily trips. An analysis by mode or purpose shows that the differences concern mainly a lower proportion of walking, a mobility by car slightly weaker for active persons, and a participation in the activities of accompaniments or leisure activities sharply weaker.

These data are coherent with the hypothesis of an under declaration attributable to the Web as far as we know that the risks of omission of trips statement concerns especially short trips or tours in terms of distance and duration and less constraint purposes. We can however object that socio-economic differences can explain at least partially the differences of mobility, especially those concerning walking trips. Internet users have a high level of education and annual income, which lead to a strong car ownership. They are more often executives and employed persons, who work on the town centre. They spend more time out of the place of residence, reducing the possibilities of participation to less constrained activities and the use of the walking. The analysis of time and distance budgets show that web respondents have a time budget superior and a distance budget slightly lower than the ones noticed in the standard survey. According to the controversial hypothesis of Zahavi concerning the constancy of the budget time (Zahavi, 1979), it could be the large budget time of web respondents which limit their mobility to a weaker level than the one of the face-to-face respondents.
Data analysis does not allow choosing between these hypotheses. To go farther, it is necessary to develop econometric techniques, allowing to take into account socio-economic differences remaining within the populations and to identify a mode effect. Attention has been paid to data comparability issue in certain disciplines, but it has not yet received much coverage in

the field of transport. We propose that the two-stage model proposed by Heckman in 1974 should be adapted. Application to the Lyon household travel survey in which two survey media were used has revealed the presence of a selection bias. The traditional estimation of a travel equation using the ordinary least squares method for the entire population therefore ignores an important factor: the determinants of survey mode choice. On the other hand, taking account of this selection bias provides a means of estimating unbiased relationships. It would be beneficial to develop this type of approach in the area of travel surveys in order, ultimately, to propose methods for combining samples obtained with different methodologies in order to increase the statistical representativeness of the surveys and reduce the non response problem.

This research thus represents an important step for modelling and surveying, although Net surfers cannot be regarded today as representative of the French population. A direction for further research on the 2006 Lyon household travel survey is to estimate the difference in the travel behaviours of the two groups by evaluating the difference between the average number of trips that the Internet respondents would have reported in a face-to-face interview and the number of trips they actually reported on the Internet. It would then be possible to allocate "correcting coefficients" to the Internet respondents with a view to integrating this sample with the data base of face-to-face respondents in order to correct the entire resulting data base to make it representative of the population.

Acknowledgments

This research has been granted by French Ministry of Transport (DRAST), SYTRAL (Lyon Public transport Authority) and Région Rhône-Alpes (Rhône-Alpes Regional council). Fieldwork has been realised by SOFRECO, AFFIMETRIE and CONVERSOFT. We shall thank Marie-Odile GASCON from CERTU for her remarks on the web questionnaire and all interviewees of the two test phases for their comments.

References

Alsnih R. (2004), New technology and travel surveys: the way forward, Working Paper ITS-WP-04-11, Institute of Transport Studies, University of Sydney, April 2004, Australia.

Ampt E.S. (1997), Response Rates - Do they matter? in Les enquêtes de déplacements urbains : mesurer le présent, simuler le futur, Bonnel P., Chapleau R., Lee-Gosselin M. & Raux C. (Eds), collection Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon, pp. 115-125.

Atrostic B.K. & Burt G. (1999), *Household non-reponse: what we have learned and a framework for the future*, in statistical Policy working paper 28, Federal Committee on Statistical methodology, Office of Management and Budget, Washington, pp. 153-180.

Bayart C., Bonnel P. & Morency C. (2009), Survey mode integration and data fusion: methods and challenges, in Bonnel, P., Lee Gosselin, M., Zmud, J. & Madre J.-L. (Eds), Transport Survey Methods: Keeping up with a Changing World, Emerald press, pp: 587-611.

Bayart, C. & Bonnel, P. (2008), *Enquête web auprès des non-répondants de l'enquête ménages déplacements de Lyon 2005-2006*, Rapport pour le PREDIT (Programme de Recherche et de Développement pour l'Innovation et la Technologie dans les transports terrestres), Laboratoire d'Economie des Transports, Lyon, 262p.

Certu (2000), L'enquête ménages déplacements « standard Certu », Collections du CERTU, éditions du CERTU, Lyon, 204p.

Certu (1998), L'enquête ménage déplacements « méthode standard », Collections du CERTU, éditions du CERTU, Lyon, 295p.

Dillman D.A. & Bowker D.K. (2001), The Web Questionnaire Challenge to Survey Methodologists, in: Dimensions of Internet Science, Edited by Ulf-Dietrich Reips & M. Bosnjak, Pabst Science Publishers, Lengerich, Germany.

Heckman, J. (1979), Sample selection bias as a specification error, Econometrica, Vol. 47, n°1, pp. 153-161.

Heckman, J. (1990), Varieties of selection bias, The American Economic Review, Vol. 80, n°2. Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, pp. 313-318.

Lozar Manfreda K. and V. Vehovar (2002), Do mail and web surveys provide same results?, In: Development in social science methodology (A. Ferligoj and A. Mrvar, ed.), Ljubljana, pp. 149-169.

Maddala, G.S. (1986), Limited dependant and qualitative variables in econometrics. Cambridge University Press, London.

http://firstmonday.org/issues/issue7_12/gunn/index.html (visited on 2005/10/26)

Morris J. & Adler T. (2003), Mixed mode survey, in Transport Survey Quality and Innovation, Stopher P.R. & Jones P.M. (Eds), Pergamon, Oxford, pp. 239-252.

Mroz, T.A. (1987), The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions, Econometrica, Vol. 55, n°4, pp. 765-799.

Myles R. & Tibert T. (1998), Internet surveys: do they work?, Institute for social research, Newsletter, Vol. 13, n° 1, Winter 1998, York University, Canada.

http://www.math.yorku.ca/ISR/newsletter/internet_surveys.htm (visited on 2005/10/18)

Nakosteen, R.A. and M.A. Zimmer (1980), Migration and income: the question of self-selection, Southern Economic Journal, n°46, pp. 840-851.

RESOURCE SYSTEM GROUP INC. (2002), Documentation for SBIR Phase II Final Report: Computer-Based Intelligent Travel Survey System, DTRS57-00-C-10030, Prepared for the FHWA, October 2002.

Richardson A.J. (2000), Behavioural Mechanisms of Non-Response in Mailback Travel Surveys, 79th Transportation Research Board, Washington, 18p.

Richardson A.J. & Ampt E. (1993), *The Victoria Integrated travel, activities and land-use toolkit*, VITAL working paper VWP93/1, Transport Research Centre, University of Melbourne.

Roy, A.D. (1951), Some thoughts on the distribution of earnings, Oxford Economic Papers 3, pp. 135–146

Stopher P.R. & Jones P.M. (2003), Summary and future directions, in Transport Survey Quality and Innovation, Stopher P.R. & Jones P.M. (Eds), Pergamon, Oxford, pp. 635-646.

Stopher, P.R. (1998), A review of separate and joint strategies for the use of data on revealed and stated choices, Transportation, Vol. 25, pp. 187-205

Zahavi, Y. (1979), The 'UMOT' Project, Report prepared for the U.S. Department of Transportation and the Ministry of Transport of Federal Republic Of Germany, 267 p.