A COMPARATIVE STUDY OF WEB- AND PAPER-BASED TRAVEL BEHAVIOUR SURVEYS

Claude Weis, Andreas Frei and Kay W. Axhausen
Institute for Transport Planning and Systems (IVT)
Swiss Federal Institute of Technology, Zurich

Thomas Haupt and Bernhard Fell
PTV AG, Karlsruhe

1. CONTEXT

The paper describes the experiences made with a series of online and pen-and-paper travel behaviour surveys of the users of Zurich's university campus area. As of today, 50'000 people are working and studying at the Swiss Federal Institute of Technology (ETH Zurich), the University of Zurich and the University Hospital. The 3 institutions plan to build new facilities totalling a combined 150'000 square meters of gross floor area, leading to an estimated 20% increase in traffic flows.

As a basis for the construction of future usage scenarios, an assessment of their effects on the perimeter's transportation system and an evaluation of measures required to meet the increased demand, the Municipality of Zurich asked the institutions to develop a mobility plan. The first step of the project, which the Institute for Transport Planning and Systems (IVT) was contracted to carry out, was the assessment of existing demand by travel surveys. The project results are published in Weidmann *et al.* (2008).

2. KEY DATA OF THE SURVEY

2.1 Study area and time line

The university campus area in the centre of Zurich, which is the main study perimeter, is displayed in Figure 1. It encompasses the headquarters of the 3 institutions under study, namely ETH Zurich, the University of Zurich, and the University Hospital. The city centre is to the west of the area shown.

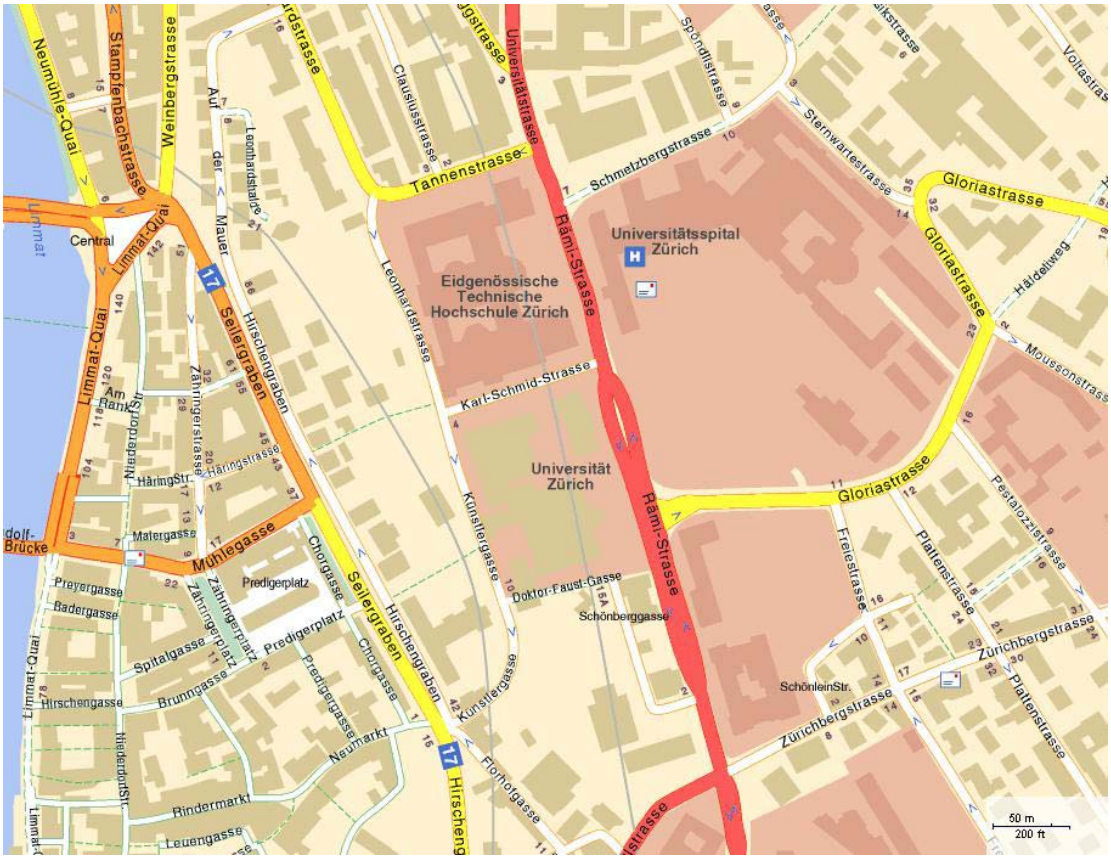

Figure 1 Study area

The users of the perimeter causing the traffic flows in and out of it can be roughly subdivided into 2 categories: members of the respective institutions (commuting to the area on a regular basis) and external visitors (with unique or periodical trips to the perimeter).

The detailed subdivision of user groups for the various institutions is displayed in Table 1. For ETH and the University, these are students, PhD students, research personnel, professors and other employees. As for the University Hospital, the employees' categorisation could not be further broken down. Here, patients and visitors were separately surveyed.

	ETH	University	University Hospital
Students	x	x	
PhD students	x	x	
Research personnel	x	x	
Professors	x	x	
Other employees	x	x	x
Patients			x
Visitors			x

Table 1 Respondent groups

The surveys of the different user groups were conducted in the time period between November 2006 and December 2007. The time line is displayed in Figure 2.

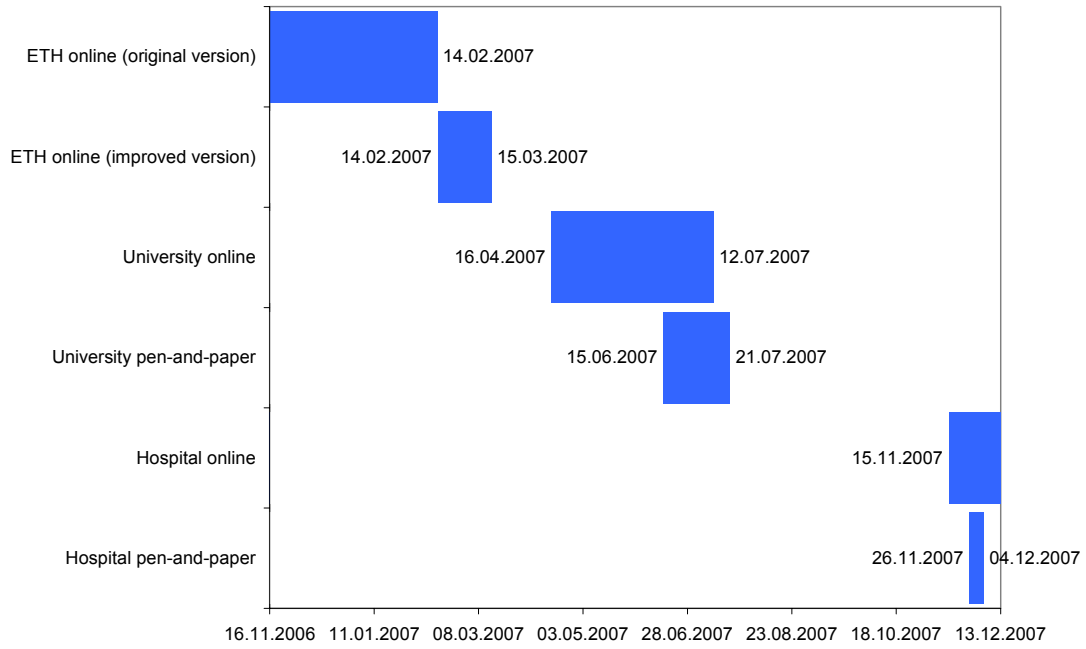

Figure 2 Time line of online and pen-and-paper surveys

The online surveys at ETH started in November 2006. Over the course of this first survey period, various problems concerning user guidance and technical compatibilities were experienced. Therefore, an improved version of the questionnaire application was introduced after approximately three quarters of the ETH surveys. This version was kept for the online surveys at both other institutions, albeit with slight methodological changes. The surveys at the University and the University Hospital were conducted in the spring and winter of 2007, respectively. In parallel, pen-and-paper questionnaires were dispatched to those users who could not be reached online.

2.2 Methodology and sampling

The various user groups mentioned above were surveyed using different methodologies that differed by two main criteria:

- Platform: part of the surveys were conducted via a web application (with a unique URL to the questionnaire being e-mailed to each respondent), part via traditional pen-and-paper questionnaire.
- Scope: part of the respondents were asked to fill in (different variants of) a complete travel diary, others were only asked about their trip to and from the study area.

The allocation of the various survey methods to the user groups as well as the corresponding sample sizes are displayed in Table 2. The slight difference in the subdivisions between ETH and the University results from the a-priori knowledge of the sample distributions. In fact, the University allocates students and PhD students to the same administrative group, so that further breakdown was not possible in advance.

	ETH		University		University Hospital		Total
	Online	Paper	Online	Paper	Online	Paper	
Students	9’060						9’060
PhD students / Research personnel	4’955						4’955
Professors	319						319
Other employees	1’955			374	4’619		6’948
Students / PhD students			7’750				7’750
Research personnel / Professors			3’888				3’888
Patients						1’183	1’183
Visitors						432	432
Total	16’289		11’638	374	4’619	1’615	34’535

Table 2 Sample sizes

The choice of the different survey methods arose one the one hand from the needs and capabilities of the various user groups. For data protection reasons, the visitors and patients of the University Hospital could not be surveyed by e-mail or mail questionnaires; thus, they had to be surveyed by questionnaires distributed on-site. For the University, the list of available e-mail addresses was not complete, so that part of the sample had to be sent paper questionnaires. On the other hand though, the large sample sizes were seen as an opportunity of testing the methodologies and documenting the differences in response behaviour.

The configuration of the various survey instruments is displayed in Table 3. The numbers indicate the ordering of the questionnaire parts.

Sociodemo-graphics	Week retro-spective	Diary			Individual trip
		Activity based	Trip based	Stage based	
ETH, online (original)					
1	2	3			
1	2		3		
1	2			3	
ETH, online (improved user guidance and navigation)					
1	2	3			
1	2		3		
1	2			3	
University, online					
1	2			3	
1	3			2	
University Hospital, online					
1	2			3	
1	3			2	
University, pen-and-paper					
1	2			3	
University Hospital, pen-and-paper					
2					1

Numbers indicate the ordering of the questionnaire parts.

Table 3 Survey instruments

All survey instruments share a questionnaire on the respondents' sociodemographic characteristics, encompassing personal data such as gender, year of birth and home address, as well as several questions on mobility tool ownership (driving license, car ownership, and public transport season tickets). The sociodemographics questionnaire came after an introduction page, where the respondents were asked if they had undertaken any trips in the perimeter on the reporting date. Conditional on the response, the diary questionnaire would be displayed or not.

In the online survey at ETH, the sociodemographics questionnaire was followed by a retrospective of trips to the perimeter for the week prior to a randomly assigned reporting date. The arrival and departure times as well as the modes for inbound and outbound trips were surveyed.

The final part of the online survey consisted of a travel diary for the reporting day, each respondent being randomly assigned one of three diary types – activity, trip, or stage based. Over the course of the first survey period, several

technical problems showed up, namely compatibility issues for various web browsers resp. their built-in pop-up window blockers. Thus, part of the respondents were not able to access the survey platform. The issue was somehow attenuated by the implementation of a warning message prompting the users to turn their pop-up blockers off. However, it was not possible to rectify it completely, so that a considerable amount of respondents were kept from accessing the survey throughout the survey period, resulting in lower than expected response rates.

Another issue turned up during the analysis of the first datasets. In fact, a large share of the respondents had only reported one item in the diary. Experiences from prior studies as well as a-priori expectations suggest a by far higher number, resp. at least two trips (one inbound and one outbound) for a mobile day. The reason for the obvious underreporting was assumed to be the survey instrument's user guidance. The labelling of the navigation bar was somewhat unclear, leading to respondents unwittingly being forwarded to the end of the survey instead of the next diary item.

In an effort to counteract this issue, an improved version of the questionnaire software was implemented where at the end of each diary item, the respondents would explicitly be asked whether or not they had undertaken any further activities to be reported. This improvement led to a slight increase of reported items. However, the number was still well below the expected average mobility, which leads to the assumption that the underreporting was due in part, but not exclusively, to the questionnaire design. Attrition certainly played its role, as some respondents may have considered providing detailed data for their cut-off date redundant after having responded to the week retrospective.

In order to assess this attrition effect, slight changes were made to the survey platform before the follow-up survey at the University. As a variant, the order of the week retrospective and diary was inverted for approximately half of the respondents. Furthermore, as the stage based diary had proven to be the most effective during the ETH survey (as far as correctly reported daily routines were concerned), it was kept as the sole diary version. The variant where the diary was put before the week retrospective rather than after it, yielded a slight, although not significant increase in reported mobility, as will be detailed in the following sections along with other analyses of response behaviour and data quality.

3. ANALYSIS OF RESPONSE BEHAVIOUR

3.1 Response rates

The time line of response rates (as the share of respondents having opened and filled the first page of the questionnaire) is shown in Figure 3.

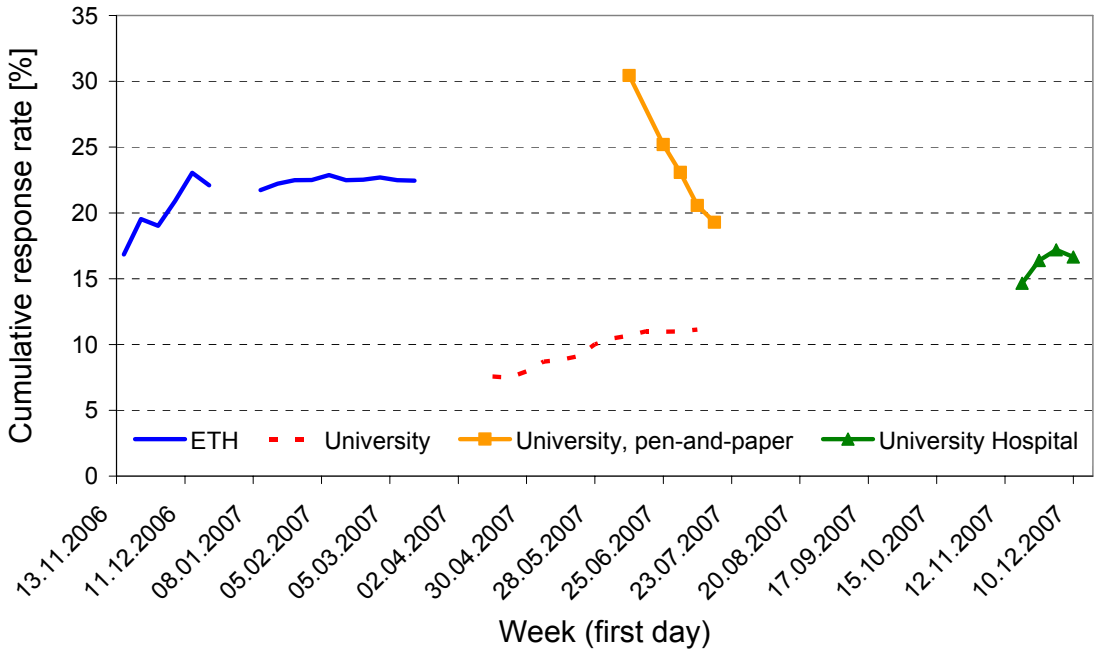

Figure 3 Time line of response rates

For the ETH survey, response rates are quite low throughout the first two weeks of the survey period. Three weeks after the initial reporting date, a new date was assigned and a reminder e-mail sent to those who had not yet opened the questionnaire. The increase of the response rates after the third week of the survey period may be explained by the effect of this reminder. The survey was paused over the Christmas holidays (last week of December and first week of January). In the weeks right before and after the holidays, response rates were slightly lower than on average. Overall, the response rate for the ETH survey was at 22.4 percent.

The same pattern of response rates before and after the dispatch of the reminder e-mails can be seen for the University survey. However, response kept rising even afterwards and settled down at a stable level only in the last third of the survey period. This effect might be due to respondents spreading the word about the survey and thus motivating their colleagues to participate in it. The overall response rate for the University online survey is at rather disappointing 11.1 percent.

As for the pen-and-paper survey, response rates during the first week were quite high, promising to remain constant throughout the survey period. However, this high response rate relates to a relatively low number of dispatched questionnaires during the pre-test period and was continuously reduced until reaching an average 19.3 percent.

Due to the considerably smaller sample, some delays before the start of the survey and the aspiration to have it completed before Christmas, the survey period at the University Hospital was shortened to about a month. Still, the reminder effect can be seen here, although it is somewhat attenuated by the effect of lower response rates during the last week before the holidays, as has been mentioned above.

In the following, response behaviour for the various respondent types will be analysed in more detail. The response rates for the user groups as they were known a-priori is shown in Table 4. The figures relate to the respondents having answered the sociodemographics questionnaire, hence they are lower than the general response rates in Figure 3 (due to respondents dropping out over the course of the survey, cf. Figure 4). In general, response rates for the pen-and-paper questionnaires are higher than those for the online surveys, pointing to a lower acceptance of the internet as a survey tool. This may be due to security concerns (data protection) or simply to the respondents' technical experience resp. to the compatibility issues mentioned above.

	ETH		University		University Hospital	
	Online	Paper	Online	Paper	Online	Paper
Students	16.5					
PhD students / Research personnel	14.2					
Professors	76.2					
Other employees	27.2			16.0	15.3	
Students / PhD students			9.6			
Research personnel / Professors			5.9			
Patients						30.2
Visitors						14.1
Total	18.3		8.4	16.0	15.3	25.9
Number of responses	2'979		975	60	707	418

Table 4 Response rates by respondent group

The figures show no clear trend of response behaviour for the various user groups. Response rates for professors are much higher than those for the other groups – end especially students – for ETH. The exact opposite holds for the University – here, students and PhD students have higher response rates than the academic personnel. As for the University Hospital, no further statement can be made on the groups' response behaviour, as the a-priori sample distribution is unknown. However, the near doubling of response rates in the patients' pen-and-paper survey compared to the visitors' is striking.

The a-priori response burden for the different survey instruments was determined according to the scheme detailed in Axhausen (2007). The methodology assigns weighted scores to certain question types and sums them up to calculate the overall response burden. The plot of the response rates for the various surveys against the ex-ante response burden is displayed in Figure 4. The response rates in the point of origin are the above mentioned rates for opening the questionnaire, resp. answering the introduction page. From this point on, decreasing response rates – which are plotted against the cumulative ex-ante response burden – can be seen for each subsequent questionnaire part. The decreasing rates indicate the share of respondents dropping out at each stage of the survey. The steepest slopes can be seen for the sociodemographics questionnaire as well as for the diary. This indicates a lack of willingness to provide personal information (especially home address and income data) in a web questionnaire on the one hand and a certain weariness of responding to the diary after completion of the week retrospective on the other hand, as has been mentioned before. The week retrospective seems to have been well accepted.

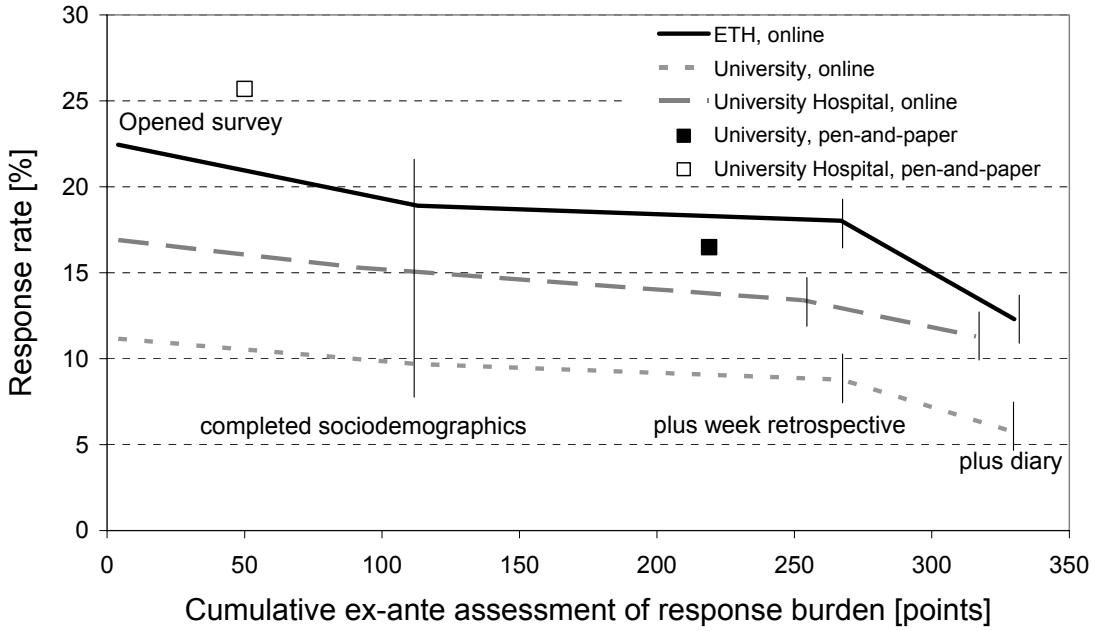

Figure 4 Response rates by ex-ante assessment of response burden

The comparison of response burden and response rates for the surveys under study in the context of other surveys conducted at the Institute for Transport Planning and Systems is shown in Figure 5. To maintain comparability, only

the pen-and-paper surveys are displayed here. It is apparent that the response rates for the University and Hospital surveys are lower than those that could have been expected from the a-priori assessment. The expectation would be around the values for the surveys without prior recruitment and without a motivation call. The link between the patients and the hospital was too weak to overcome the response burden.

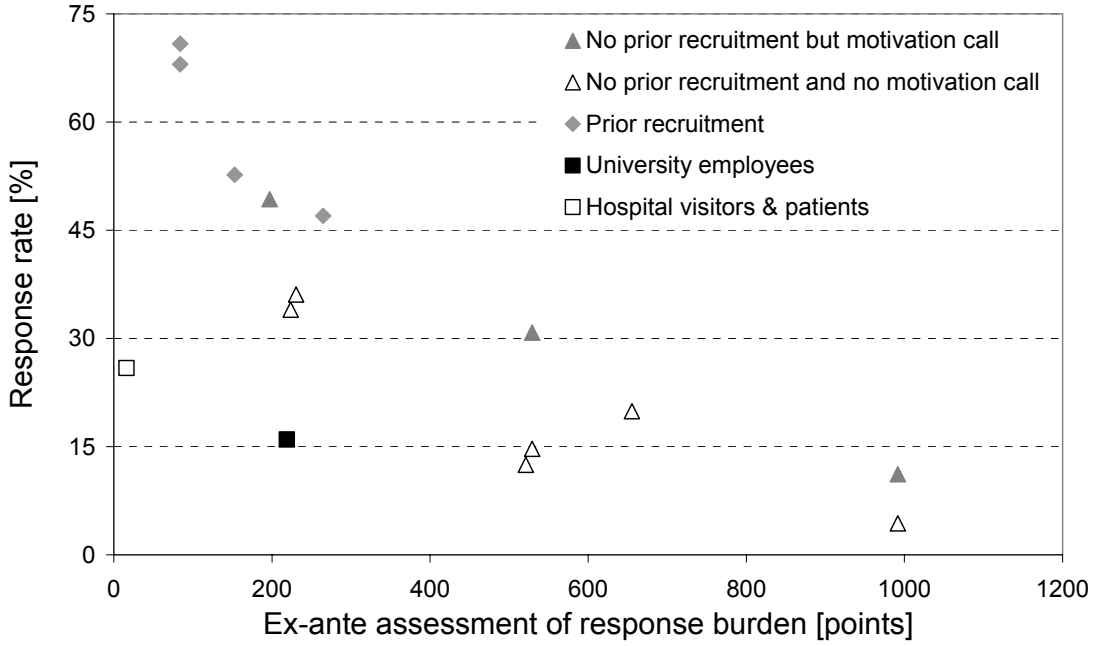

Figure 5 Response rates compared to other studies at IVT

3.2 Data quality

Apart from the willingness to participate in the survey and the resulting response rates, another interesting aspect to be considered is the quality of the data the survey yield. Here, the different diary variants will be compared, the criterion applied being the number of trips reported by the respondents. The distribution for the various diaries is displayed in Table 5.

Two findings are central for the ETH surveys. On the one hand, the number of reported rips for the activity based diary is considerably lower than that for the other two variants. The explanation for this a-priori surprising effect is that the questionnaire design did not make immediately clear to the respondents that as an activity subtype, trips ought to be separately reported.

On the other hand, the effect of improving the user guidance by implementing the explicit question whether an additional trip had been undertaken considerably increased trip reporting for the trip and stage based variants. It is still below the expected average, however this underreporting is likely due to attrition effects, and not to the survey design.

By far the highest number of reported trips was yielded by the stage based pen-and-paper questionnaires. It may be assumed that the respondents take the questionnaire along as they undergo their day and fill it in on-the-fly. Thus,

more trips are reported than for the online survey where everything has to be remembered and reported at the end of the day. Also, the attrition effects do not show for the pen-and-paper questionnaire, as incomplete diaries are likely not to be sent back at all, and thus do not show up in the statistics as opposed to the online surveys.

Socio-demographics	Week retro-spective	Diary			Number of trips
		Activity based	Trip based	Stage based	
ETH, online (original)					
1	2	3			0.69
1	2		3		1.31
1	2			3	1.33
ETH, online (improved user guidance and navigation)					
1	2	3			0.83
1	2		3		1.93
1	2			3	1.96
University, online					
1	2			3	1.97
1	3			2	2.05
University Hospital, online					
1	2			3	1.79
1	3			2	1.71
University, pen-and-paper					
1	2			3	3.12

Table 5 Comparison of diary variants – average number of trips

A further indicator for the correctness of the reported travel behaviour is the destination of the last recorded trip. If the diary is complete, the final destination should in most cases be the home location. As Figure 6 shows, this is not always the case for the datasets under study. For the ETH online survey, where lots of “one-trip-days” were reported, this unique trip often ends at the work place. This is a clear indicator for the fact that the diaries were not correctly completed, but rather only the first trip to work reported and the survey aborted afterwards, often inadvertently. The improved user guidance brings a clear improvement, which also shows in the follow-up surveys at the University and the University Hospital.

Inverting the order of the week retrospective vs. the diary did not yield a significant effect. In the pen-and-paper surveys, the reported daily routines seem to be correct, at least as far as the final destinations are concerned.

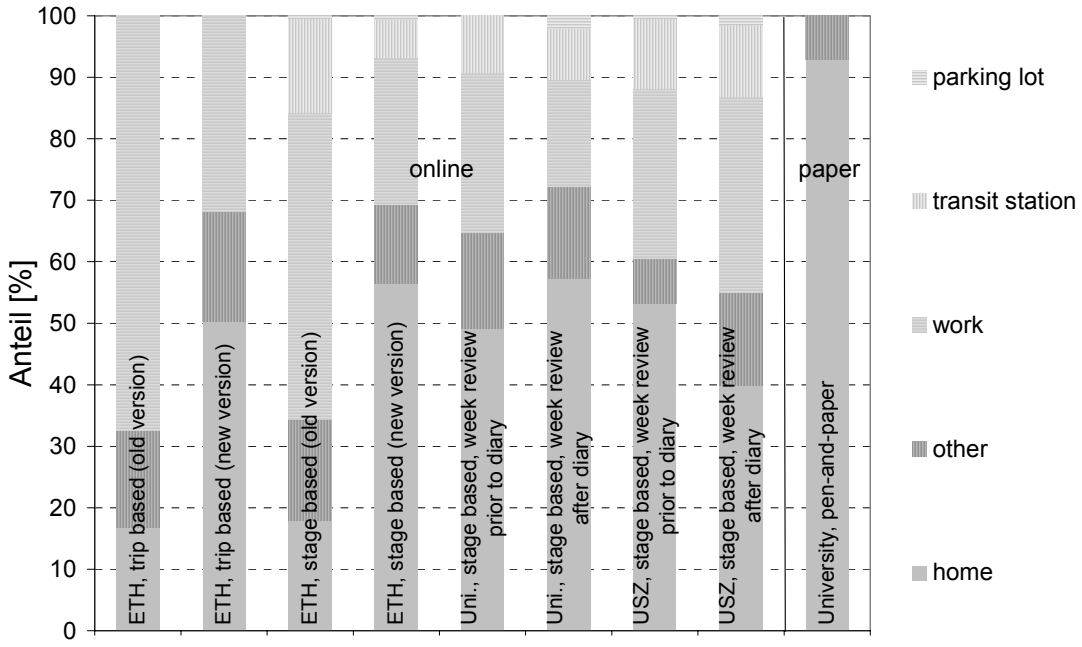

Figure 6 Comparison of variants – destination of last recorded trip

A plausibility control similar to the one above is one concerning the arrival times of the last reported trip. For a mobile work day, one would expect it to be some time between noon and midnight. Here also, there is an apparent quality difference between the original and the improved version of the online questionnaire, as is shown in Figure 7. For the original version, a majority the reported days end before 9:00 in the morning and, as mentioned above, at the work location. The improved user guidance leads to the last reported trips mostly ending during the late afternoon and thus leads to far more plausible daily routines.

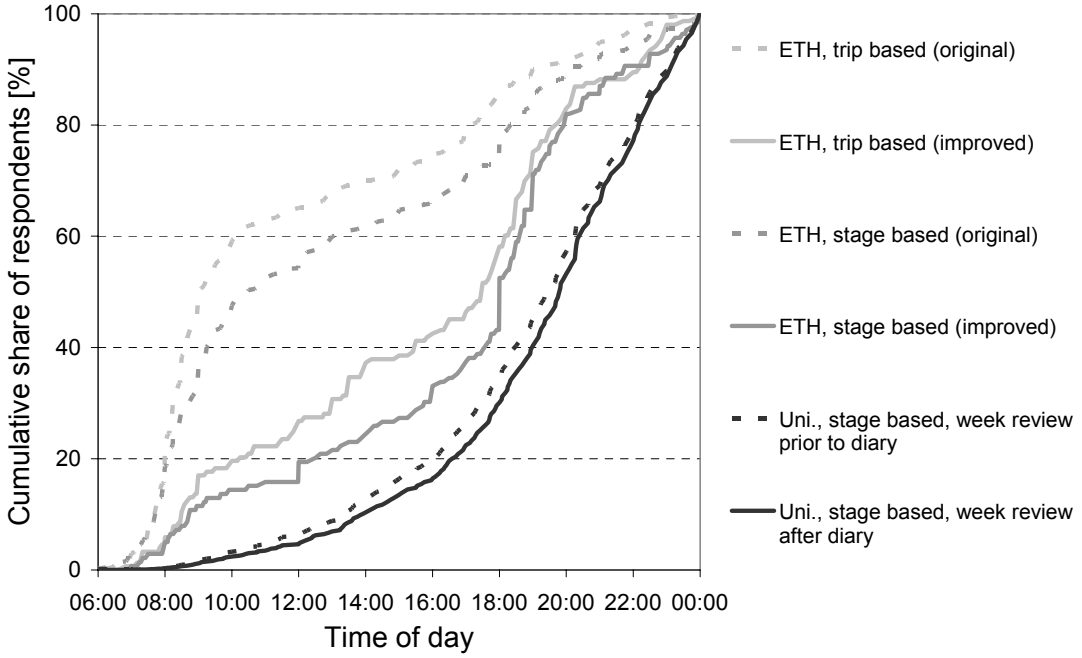

Figure 7 Comparison of variants – arrival time of last recorded trip

4. MODELLING OF RESPONSE BEHAVIOUR

4.1 One-way analysis of Variance

This section focuses on the search for a statistical affirmation of the qualitative trends established so far. Therefore, one-way analyses of variance for the number of reported trips according to various criteria were first conducted in order to determine the significance of the affects for the various survey instruments and user groups. The F-statistics and significance levels for the various indicators are displayed in Table 6.

The differences in the dependent variable are practically only significant for the subdivisions of the various survey instruments. The sociodemographic characteristics and the mobility tool ownership variables, which one would normally expect travel behaviour and inherently the number of trips undertaken by an individual, are mostly not significant. This again points to undesired effects of the survey design on response behaviour.

Criterion	F	Sig.
Survey instrument: online / pen-and-paper diary	100.4	0.000
Online: original / improved version	262.7	0.000
Online: stage / trip / activity based diary	190.2	0.000
Online ordering of retrospective and diary	95.0	0.000
Institution: ETH / University / Hospital	129.3	0.000
User group: students / professors / other	2.9	0.021
Gender: male/ female	9.1	0.003
Marital status: single / married / divorced / widowed	0.8	0.504
Car driving license: yes / no	0.1	0.918
Own car: yes / no	0.1	0.807
General public transport abonnement: yes / no	4.5	0.033

Table 6 Analysis of variance for number of reported trips

The effects yielded by the analysis of variance for the online surveys shall now be assessed in more detail by the estimation simultaneous models. Various model approaches have been applied.

4.2 Poisson Regression

The *Poisson regression* is part of the *generalized linear models (GLM)* family, which are a generalized form of the classical linear regression models. In *GLM*, the relation between independent and dependent variables is given by specific *link functions*.
GLM allow for the use of non-normal – and, more specifically, discrete – distributions for the dependent variables. They consist of three components:

- a distribution function, f, for the independent variable, Y;
- a linear predictor, $\eta = \beta X$ and
- a parametric link component g, such that $E(Y) = \mu = g^{-1}(\eta)$.

The *Poisson regression* uses the *Poisson distribution* for the independent variable. Thus, it is appropriate for modelling independent variables that take on only non-negative integer values, such as count data. The same holds for the number of reported trips to be modelled here. The *link function*, determining the relation between the predictor variable and the mean of the distribution function, has been set to the natural logarithm. The scale parameter, which is the ratio of the mean and the standard deviation of the distribution function and is normally restricted to one, has here been estimated as an additional parameter, undoing the assumption of the mean and standard deviation being equal. This phenomenon, called *dispersion*, was confirmed by the descriptive data analysis.

Table 7 shows the results of the *Kolmogorov-Smirnov test*, a non-parametric test of sample distributions against any distribution function showing which distribution is the most appropriate for modelling.

The test value is the smallest for the *Poisson distribution*, indicating that it is closest to the real sample distribution.

Distribution	Kolmogorov-Smirnov Z
Uniform	37.51
Normal	17.36
Exponential	17.11
Poisson	5.35

Table 7 Results of Kolmogorov-Smirnov test

The estimation results for the *Poisson regression* are displayed in Table 8. The same effects can be seen as for the descriptive analysis in section 3 and in the analysis of variance: the number of reported trips is mainly influenced by response behaviour as a reaction to the survey instruments design, and not by the real travel behaviour as it would be induced by the respondents' sociodemographic characteristics. The only significant parameters are those for the survey instrument categorization.

Parameter		β	t
Intercept		0.793	2.05
Questionnaire version	Original	-0.336	-6.88
	Improved	-	
Diary version	Stage based	0.019	0.37
	Activity based	-0.686	-11.49
	Trip based	-	
Order	Week retrospective - diary	-0.015	-0.26
	Diary - week retrospective	-	
Institution	ETH	0.075	0.80
	University	0.188	2.10
	University Hospital	-	
Respondent group	Students	-0.086	-0.71
	PhD students	0.050	0.42
	Professors	-0.003	-0.02
	Employees	-0.013	-0.14
	Academics (Hospital)	-	
Gender	Male	-0.005	-0.14
	Female	-	
Marital status	Single	-0.163	-0.44
	Married	-0.298	-0.80
	Living apart	-0.461	-1.08
	Divorced	-0.245	-0.64
	Widowed	-	
Car driving license	No	-0.013	-0.18
	Yes	-	
Own car	No	0.023	0.52
	Yes	-	
General abonnement	No	-0.061	-1.61
	Yes	-	
Household size		0.029	2.09
Scale parameter		1.079	
	Log-likelihood	-3'485.4	
	Likelihood ratio χ^2	473.1	
	Sig.	0.000	

Table 8 Results of Poisson regression

4.3 Negative Binomial Regression

One issue with the *Poisson regression* is that there is no formal procedure of testing the significance of the *dispersion*. One way of testing it though, is comparing the standard *Poisson regression* to the *negative binomial regression* (Nurosis, 2008). Here, the underlying distribution function for the dependent variables is the *negative binomial distribution*.

The *log-likelihood* values for the standard *Poisson regression* as well as for the *negative binomial regression* are displayed in Table 9. The *log-likelihood* value for the standard *Poisson regression* is larger than the one for the *negative binomial regression*. Thus, the *negative binomial regression* has not led to an improvement of the model fit.

Model	Log-likelihood
Poisson without dispersion	-3'485.4
Negative binomial	-3'993.0

Table 9 Comparison of model fit – Poisson vs. negative binomial regression

4.4 Ordinal Regression

Ordinal regression is applied for modelling ordered categorical variables (Long, 1997). The observed variable is considered as a categorical representation of a latent (unobserved) variable (Fahrmeir and Tutz, 2001).

The regressand is the cumulative distribution of the independent variable, here the number of reported trips. Again, the *link function* has several possible forms. In cases where the majority of values is in the lower part of the distribution, literature (Nurosis, 2008) recommends using the *negative log-log* approach, thus the *link function* is given by *–ln(-ln(p(event)))*.

Apart from the regression coefficients for the independent variables, threshold values for the latent variable are estimated using a *maximum likelihood* approach. These threshold values are used as intercepts in the calculation of the probabilities for the different categories.

The threshold values are displayed in Table 10, while Table 11 shows the estimated regression parameters. Only variables yielding significant regression coefficients in the *Poisson regression* were used in the *ordinal regression* – the survey instrument categorization as well as the institution dummy variables.

The *ordinal regression* parameter estimation confirms the trends exhibited by the previous models and the analysis of variance. The *log-likelihood* value is larger than that for the *Poisson regression*, indicating that the *ordinal regression* is the more appropriate modelling framework for the data at hand.

Threshold	τ	t
0	-2.010	-16.57
1	0.436	4.11
2	1.458	13.06
3	2.181	18.02
4	3.045	21.28
5	3.918	21.07
6	4.746	18.51
7	5.559	15.12
8	6.252	12.26
9	6.946	9.72
10	7.639	7.60

Table 10 Results of ordinal regression – threshold values

Parameter		β	t
Questionnaire version	Original	-0.718	-9.02
	Improved	-	-
Diary version	Stage based	0.06	0.78
	Activity based	-2.146	-24.15
	Trip based	-	-
Order	Week retrospective - diary	-0.097	-1.20
	Diary - week retrospective	-	-
Institution	ETH	0.045	0.40
	University	0.173	2.10
	University Hospital	-	-
	Log-likelihood	-415.6	
	Likelihood ratio χ^2	1’559.4	
	Sig.	0.000	

Table 11 Results of ordinal regression – parameters and model fit

The comparison of the real number of reported trips distribution vs. the one calculated from the model parameters is displayed in Figure 8. As can be seen, the model reproduces the real distribution quite well.

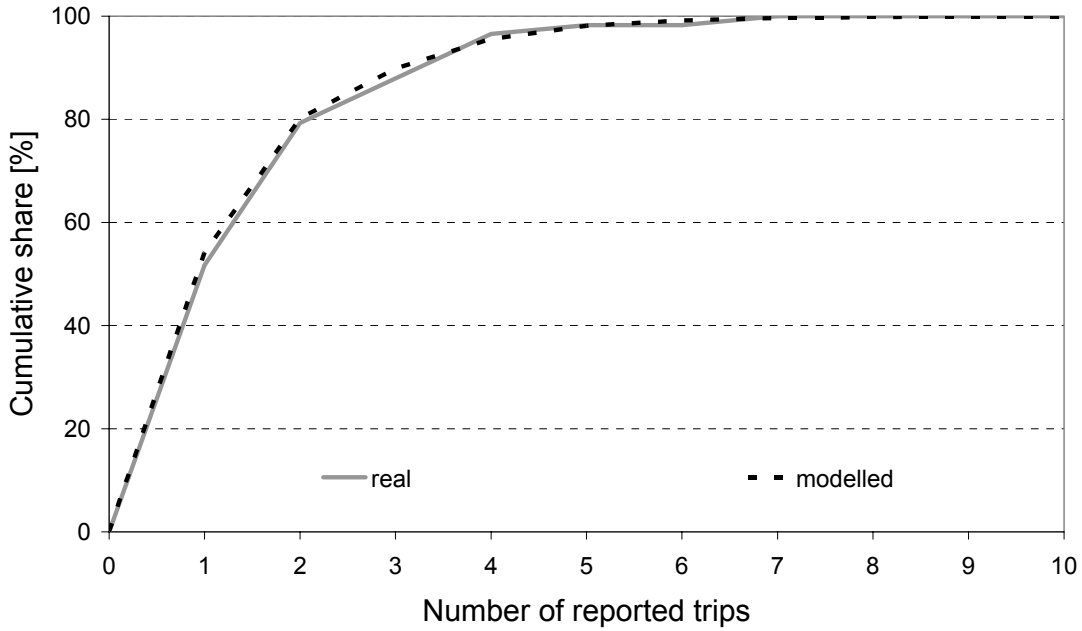

Figure 8 Real vs. modelled distribution of number of reported trips

5. SUMMARY AND CONCLUSIONS

The travel diary surveys conducted in Zurich's high school campus area and presented in this paper yielded a number of interesting findings. A large part of the interviewees' response behaviour can be explained by the design of the survey instruments. On the one hand, a classic pen-and-paper questionnaire, sent out by mail, appears to be generally well accepted and yield response rates and data consistent with what one would have expected. On the other hand, the continuous improvements made to the compatibility and the user guidance of the online survey tool led to the reporting of more realistic travel behaviour characteristics. This indicates that, especially when conducting such a survey by online tools, special care has to be taken to provide the respondents with a reliable and user friendly platform.

The advantages of such online surveys as compared to pen-and-paper diaries are obvious. The administrative expenditures are lower, and a large sample can be reached within relatively short time and, once that the platform has been set up, at quite a low cost. These advantages can be exploited, given that appropriate caution is applied. However, the data yielded by the online surveys are overall less reliable than those coming from the classic questionnaires.

The lesser acceptance of the internet as a survey tool may be due to security concerns, the technical capabilities of the respondents or comfort issues – filling a pen-and-paper questionnaire on-the-fly may be less burdensome than reporting a complete day (or week) at once.

6. BIBLIOGRAPHY

Axhausen, K.W. (2007) Predicting response rate: a natural experiment, **Note**, Institute for Transport Planning and Systems, ETH Zurich, Zurich.

Fahrmeir, L. and Tutz, G. (2001) *Multivariate Statistical Modelling Based on General Linear Models*, 2nd edition, Springer, New York.

Long, J.S. (1997) *Regression Models for Categorical and Limited Dependent Variables*, Sage, Thousand Oaks.

Nurosis, M. (2008) *SPSS 15.0 advanced statistical procedures companion*, Prentice Hall, Upper Saddle River.

Weidmann, U., Axhausen, K.W., Spacek, P., Alt, B., Anderhub, G., Dorbritz, R., Frei, A., Laube, M., Scherer, M. and Weis, C. (2008) Mobilitätsplan Hochschulgebiet Zürich, Hauptbericht, Institute for Transport Planning and Systems, ETH Zurich, Zurich.

