Master Thesis

University of Basel

Faculty of Business and Economics

Using a Nudge to Reduce Mobility Externalities: Evidence from a RCT with GPS Tracking in Switzerland

Author
Laura Schwab
(2017-055-310)

Supervisor
Prof. Dr. Beat Hintermann

Master of Science in Business and Economics
Major: Markets and Public Policy

14 June 2022

Abstract

This is the first study to conduct a nudging-related RCT in the context of mobility. The study is designed as an RCT with a control group to examine the effect of recurring nudging reports including social comparisons on total external costs. 777 participants from the German- and French-speaking parts of Switzerland were tracking themselves using the GPS-based app "Catch-my-Day". Nudging reduces external climate costs by 3.5 Swiss cents per day in absolute, or 5.7 percent per day in relative terms. Total external costs including health, congestion and climate externalities are reduced by 2.0 percent, but not to a significant extent. Reports containing positive feedback cause reductions in external costs, whereas negative feedback leads to a boomerang effect with participants generating even more external costs. Treatment effects diminish over time, resulting in no significant long-term effects. The average participant reveals a mean WTP between 0.62 and 2.42 Swiss Francs per nudge.

Contents

List of Figures

List of Tables

1 Introduction

Transport activities across all transport modes have steadily increased in recent years in Switzerland (Statista, 2022). The Swiss Federal Statistical Office (2018) estimates total direct costs of transport at 94,9 billion Swiss Francs in 2018 (excluding external costs). As in most countries, Swiss public transport is subsidized to a large extent, and can be compared to a natural monopoly.
Transportation behavior also creates market failures due to several types of external costs and benefits associated with different modes of transportation. Important examples are congestion, accidents, physical activity and emissions such as noise, air pollution or CO_2. To assess the welfare impacts of transportation, externalities leading to market failure need to be taken into account.

Generally, externalities can be described as external effects of economic activity of an agent on society that are not taken into consideration by the agent, i.e., do not enter his or her utility function (Verhoef, 2000). In a first-best world, any externalities are internalized by charging users the marginal external costs. In other words, charging individuals the exact amount of their marginal external costs leads users to adjust their consumption decision and thereby restore the socially optimal level of mobility (Verhoef, 2000). In reality, implementation difficulties occur. Marginal external effects of all traffic participants are rarely ever completely known. According to Verhoef (2000), one should still aim at internalizing mobility externalities as far as possible in a second-best world.
The large-scale market failure in the transport sector serves as normative motivation for policy interventions attempting to reduce several kinds of externalities. According to Delft (2019), general external transportation costs can be divided into costs related to accidents, air pollution, climate change, noise, congestion, well-to-tank emissions, habitat damage, and other external costs.
The classical “hard” approach aims at internalizing externalities either by rules and regulations (Möser and Bamberg, 2008), or by prices, e.g. using Pigouvian transport pricing (Axhausen et al., 2021). Besides, there has been rising interest in “soft” transport policy measures in recent years (Möser and Bamberg, 2008). These approaches aim to influence individual decision making by persuading people to change their perceptions and motivations (Möser and Bamberg, 2008). An example of such a “soft” policy is information-based nudging. Thaler and Sunstein (2008) define nudging as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives”.

This nudging concept is implemented in this study. The study uses the design of a randomized controlled trial (RCT) being conducted between July 2021 and February 2022. The sample consists of a total of 777 people living in the German- and French-speaking parts of Switzerland. The treatment group receives a recurring nudge in form of weekly reports including social comparison as well as information about the participant’s external costs of the current week. The study uses a control group to absorb time-varying factors which may be correlated with the treatment. This study focuses on the reduction of the most important external costs of transport, which are congestion, climate damage, and health effects. These three categories summarize externalities related to accidents, air pollution, climate change, noise and congestion. However, they leave out externalities associated with well-to-tank emissions and habitat damage (Delft, 2019).

Research in economics as well as psychology has shown behavioral interventions can be powerful tools in shaping people's behavior in a variety of domains (Andor and Fels, 2018). Andor and Fels (2018) emphasize that "non-price measures are relatively inexpensive to implement and do not interfere with people's choice sets as strongly as, for example, taxes or bans on certain products". Moreover, as Hintermann et al. (2021) state, non-financial interventions could be easier to implement than prices or taxes because of concerns of social acceptability. This serves as the primary motivation to conduct a study examining non-price interventions in the mobility sector in more detail. Since this study focuses on nudges, an overview of the existing nudging literature is presented in the following paragraphs.

The whole nudging concept started to roll with Schultz et al. (2007) and Nolan et al. (2008) emphasizing that normative social influence and particularly social norms can be powerful tools in energy conservation. Originally, the idea of using reports as recurring nudges originates from an experiment that uses social comparisons in home energy reports (HERs) to reduce energy consumption (Allcott, 2011). Allcott (2011) conducts a randomized natural field experiment with 600'000 treatment and control households across the US that receive HERs from a company called OPOWER. The HERs compare the energy use of a household to similar neighbors and provide energy conservation tips. The program shows an average reduction of 2.0%, ranging from 0.3% in the lowest decile to 6.3% in the highest decile. This paper inspired countless other researchers to conduct experiments on related topics. Hence, a large part of the existing literature on nudging deals with information-based measures to reduce energy consumption. The most relevant studies are summarized below.
Allcott and Rogers (2014) show that social comparison-based home energy reports lead to both substantial and significant energy reductions in the short term, and smaller but still significant energy reductions in the long term. Myers and Souza (2020) conduct an RCT to investigate the effect of comparison-based HERs on the energy consumption of households. The authors find almost no behavioral changes in heating demand. Löschel et al. (2020) find no significant reduction in energy consumption when implementing an energy savings application for mobile phones. In an Italian field experiment, Marangoni and Tavoni (2021) show that the provision of real-time feedback on electricity consumption significantly decreases electricity consumption by 1-2%. In their meta analysis, McKerracher and Torriti (2013) find that real-time feedback on electricity consumption provided through in-home displays reduces consumption by 3-5% on average. Delmas et al. (2013) find an average energy reduction of 7.4% in their meta-analysis of information-based energy conservation experiments.

Nudging RCTs have been applied to different topics. Ferraro and Price (2013) implement a natural field experiment with more than 100'000 households to show that social comparison messages are most effective in significantly reducing residential water demand. Wilson et al. (2017) conduct an RCT in the US to show that nudging significantly increases the uptake of targeted foods. Tyers (2018) finds no significant increase in voluntary carbon offsetting for air travel when implementing a nudging RCT intended at increasing pro-social behavior.

Sasaki et al. (2022) use an RCT in Japan with 1'595 persons to show that nudging significantly increases the number of people receiving the COVID-19 vaccine. Dai et al. (2021) find similar results for the US.

Several studies investigate the effect of informational interventions in the transport sector using a before-after comparison. Pluntke and Prabhakar (2013) implement a project that manages peak demand by incentivizing commuters to travel off-peak by using social comparisons and personalized offers. Their program effectively induces commuters to shift from peak to off-peak travel times. Bothos et al. (2014) evaluate an app that nudges users to travel environmentally friendly. In their before-after study without a control group, the authors find neither a significant effect on mode choice nor on environmental concerns. They find a significant positive effect on attitudes towards sustainable travels, though. Similarly, Carreras et al. (2012) use a before-after setting without a control group in Finland to evaluate the effect of an app and open-source platform on more sustainable mobility choices of users. Despite using a combination of goal-setting, self-monitoring and sharing features, they do not find any significant effect neither on travel behavior nor on environmental attitudes. Jariyasunant et al. (2015) conduct a before-after analysis without a control group to evaluate an app that aims at changing travel modes by providing quantitative feedback on travel behavior. The authors find a significant reduction in distance traveled, and significant psychological effects such as more awareness of environmental consequences of transport, more open-minded attitudes towards sustainable mobility, and the intention to use more sustainable transport modes. Maerivoet et al. (2012) conduct an experiment to show that when equipping vehicles with on-board units using smart technology to provide current cost information, road users adapt their behavior, resulting in lower costs of transport.

The following section summarizes the few information-based RCTs that have been carried out in the transport context. Cellina et al. (2019) evaluate the effect of a mobility tracking app that provides feedback on environmental friendliness as well as social comparisons. Their RCT consists of 52 test persons suitable for the analysis in the Swiss cantons of Zurich and Ticino. The authors conclude that the app significantly reduces CO_2 emissions and energy consumption per kilometer for regular travelers in Ticino. Rosenfield et al. (2020) find no statistically significant effect in an RCT with approximately 2'000 commuters analysing whether weekly informational emails decrease the frequency of commuting by car. Kristal and Whillans (2020) conduct an RCT to investigate whether different informational treatments increase the share of carpooling. They find no significant effect of behavioral nudges on the reduction of single-occupancy vehicle commuting. Gravert and Collentine (2021) investigate whether public transport usage can be influenced by social norms and economic incentives using a natural experiment with over 14'000 individuals. The authors don't find any effect for descriptive social norms on ridership, but they show that increasing economic incentives significantly increases uptake and long-term usage of public transport. Axhausen et al. (2021) conduct an RCT using the GPS tracking app "Catch-my-Day" with 3'700 participants in Switzerland, being randomly assigned to control, pricing and information treatment after the observation period. Participants with pricing treatment significantly reduce external costs, whereas participants in the information group also show reductions, but not to a statistically significant extent. To sum up, the existing literature provides no clear evidence of a significant effect of informational nudging in the transport area up until now.

This thesis also relates to recent literature that elicits willingness-to-pay (WTP) or willingness-to-accept, and then evaluates welfare effects of informational policy instruments such as energy-use social comparisons (Allcott and Kessler, 2019), reminders (Damgaard and Gravert, 2018), and calorie labeling (Thunström, 2019).

Undoubtedly, tools of behavioral economics are gaining increasing popularity in various research disciplines (Hummel and Maedche, 2019). This thesis adds two main contributions to the existing literature. First, to the best of my knowledge it is the first study to transfer the concept of HERs from the energy to the transport sector. Second, the study provides insight into private costs including personal benefits or costs related to receiving reports, which is gained through a revealed-preference approach that elicits individual WTP. Including private costs that people incur when being nudged is essential for drawing conclusions about the overall welfare implications of nudging.

The thesis is structured as follows. Section 2 explains the study design of the experiment. Section 3 presents the methodology needed for the regression estimates, as well as the methodology used to conduct the WTP analysis. Section 4 presents the data. Section 5 presents the results, Section 6 discusses them, and Section 7 concludes.

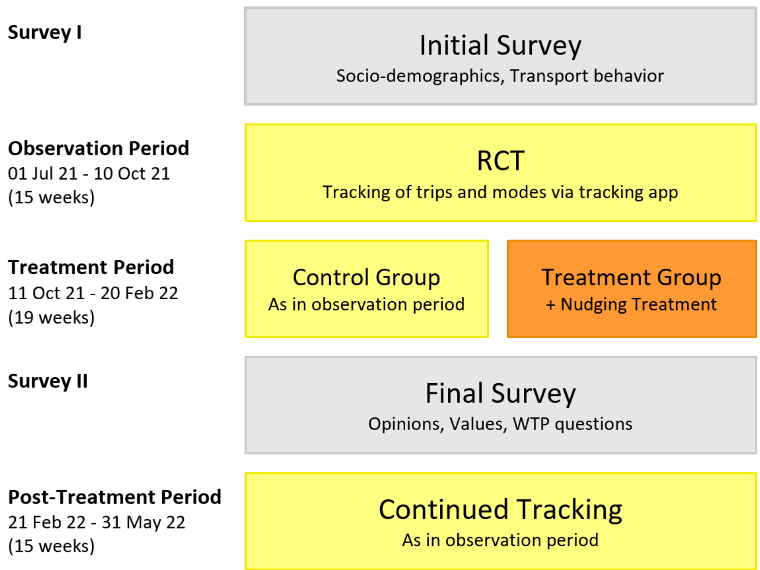
2 Experimental design

2.1 Study design

The study sample was randomly recruited from individuals living in the German- and French-speaking parts of Switzerland. Some of the participants of this study had already been recruited for the MOBIS experiment by Axhausen et al. (2021), or the MobisCovid study by Molloy et al. (2021). Besides re-inviting former study participants, about half of the participants were newly recruited via the LINK panel.[1] The opt-in design of this study allowed participants to quit tracking whenever they wanted. Participants were not promised any remuneration in return for taking part in this study. However, 10 percent of the treatment group who had filled out the final survey were paid out one of their given options in the questions about individual WTP.

Figure 2.1 provides an overview of the study design. Regardless of the different recruiting dates, each participant had to fill in an initial online survey with questions about travel behavior and socio-demographics. After the observation and treatment periods, participants were asked to fill in a survey containing questions about their opinions, values, and WTP for additional reports. After the final survey, tracking was continued to gain insights into the post-treatment mobility behavior of participants.

Figure 2.1: Study design



2.2 GPS tracking

Study participants were tracked via the smartphone tracking app “Catch-My-Day”. The Motiontag analytics platform then imputed trip stages, travel modes and activities to each movement. The Motiontag app split a day into stages (segmentation) using a machine-learning

[1] See also `https://www.link.ch/`.

algorithm. Participants were able to review and correct the automatic mode detection. The following modes were automatically assigned by the "Catch-my-Day" app: airplane, bicycle, bus, car, ferry, train, tram and walk. Users could select the following options as corrections: boat, car-sharing, e-bike, e-scooter, gondola, motorbike/scooter, Pikmi, taxi/Uber, ski, and subway. Nowadays, state-of-the-art machine learning algorithms for mode and activity detection achieve accuracy rates of over 90%, depending on the approach (Wu et al., 2016; Nikolic and Bierlaire, 2017). For more information on the GPS tracking and mode detection process implemented in this study, see Axhausen et al. (2021) and Hintermann et al. (2021).[2]

2.3 The external costs of transport

Three types of externalities are considered in the RCT: congestion, health (including noise) and climate external costs. Throughout the analysis, marginal external costs are considered. The costs of the mobility behavior are computed on the recorded daily trips using an automated data pipeline. Additionally, data collected from the online introduction survey are incorporated into the data processing to improve the imputation. The external costs are calculated analogously to Axhausen et al. (2021). For a more detailed description, please refer to them.

Costs associated with driving

To calculate the external costs of private road transport, GPS tracks are mapped to the Swiss road network using the open-source routing server GraphHopper (Karich and Schröder, 2014). Figure 2.2 displays how tracks are then combined with emission factors from the HBEFA database, and converted using the MATSim emissions module (Hülsmann et al., 2011; Kickhöfer et al., 2013). Congestion costs are measured in seconds of delay, climate externalities in grams of CO_2, and health externalities in grams of caused nitrogen oxides (NO_x) and particular matter (PM_{10}). Based on the values in Table 2.1, the externalities are converted into monetary units.

Table 2.1: Monetization of externalities

Emission	Aspect	Value	Unit
Scenario year		2019	
CO_2	Climate Costs	136.08	CHF/ton
PM_{10} Costs (Healthcare)	Rural	515,497	CHF/ton
	Urban	1,358,461	CHF/ton
NO_x	Regional	7,109	CHF/ton
VTTS	National	25.77	CHF/h [a]

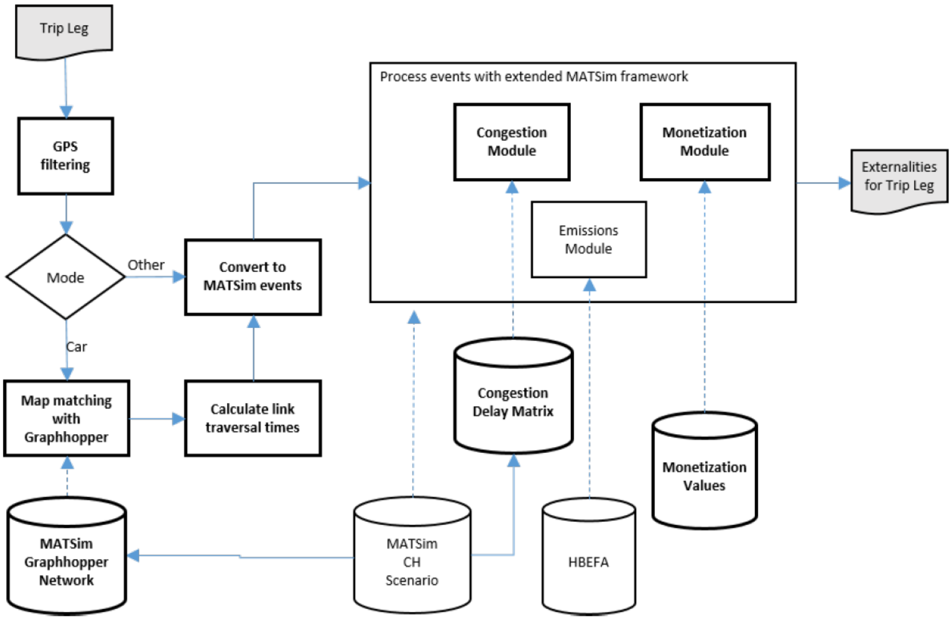
Source: Federal Roads Office - ASTRA (2017), updated for 2019; *a* scaled nominal wage rate

Costs of active and public transport

For modes other than driving, per-km values presented in Table 2.2 are applied. Health costs

[2]Most of the information in Section 2 (and also in later Sections) is based on Axhausen et al. (2021) and Hintermann et al. (2021), as this study directly builds on their study design. For practicality reasons, they are not always cited again.

Figure 2.2: MATSim-based externalities pipeline



Source: Axhausen et al. (2021)

for active modes consist of the sum of accident costs and health benefits. For bicycling, the healthcare costs due to accidents outweigh the benefits, which results in a positive amount of total externalities associated with moving by bicycle. Walking is the only mode of transport to exhibit external benefits by reducing healthcare costs. To reflect crowding in public transport, congestion externalities are included for public transport at specific times and places using a zonal peak-hour surcharge pricing scheme.

Table 2.2: Per-km monetary costs (in CHF) used in the experiment

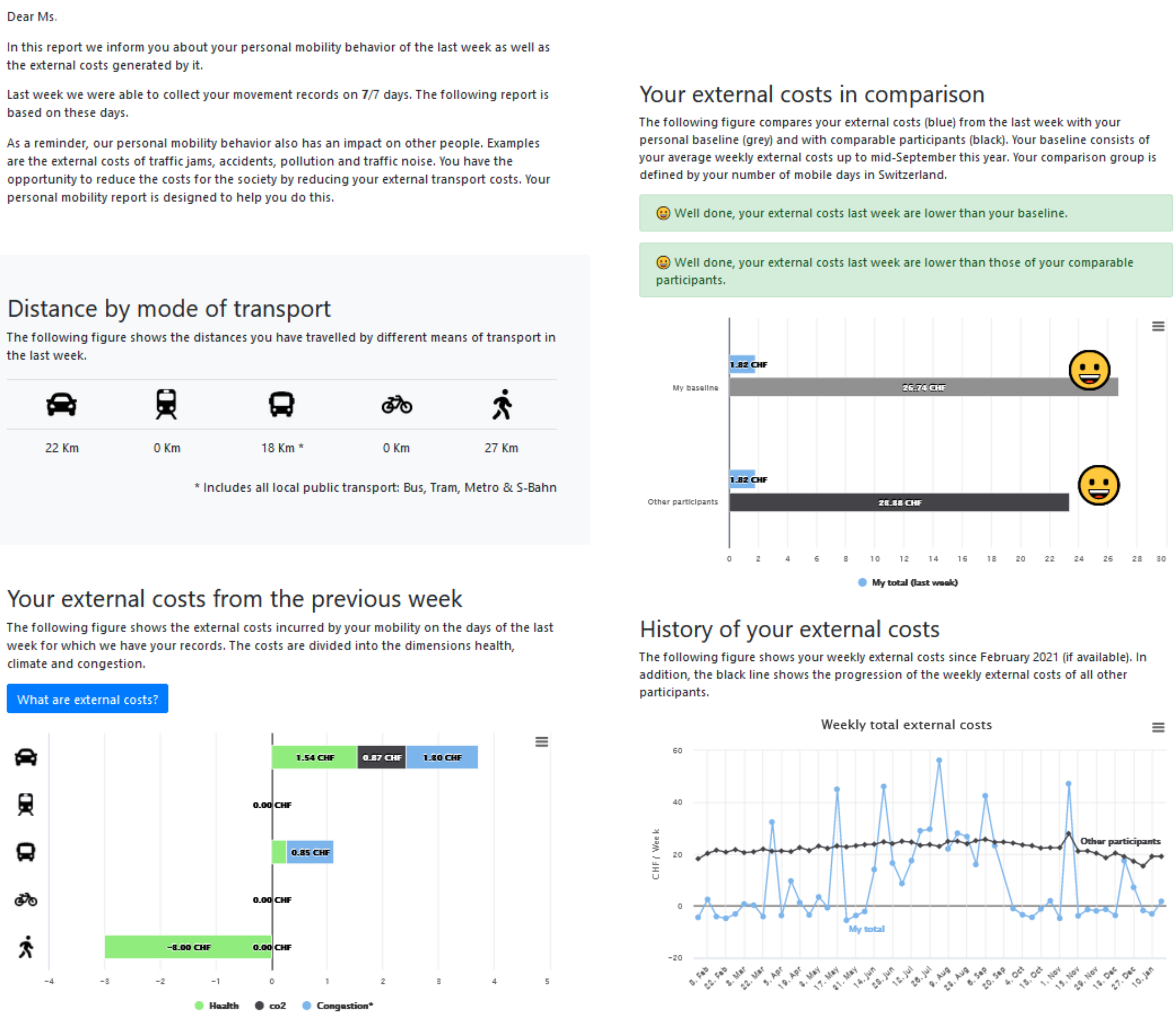
Mode	CO_2	PM_{10}	NO_x	Accidents	Noise	Health
Train	0.000066	0.0140	-	0.00066	0.0087	-
Bus	0.0144	0.0437	0.5440	0.0141	0.0257	-
Tram	-	-	-	0.0126	0.0075	-
Bicycle	-	-	-	0.257	-	-0.1870
Walk	-	-	-	0.075	-	-0.1863

Notes: Table re-displayed from Hintermann et al. (2021). CO_2 values for tram are omitted, but they would be small. Negative costs indicate an external benefit.

2.4 Treatment

The RCT consists of 15 weeks of observation for all participants, followed by a treatment period of 19 weeks in total. For the control group, tracking was continued as before during the treatment period. The assignment to the treatment and control group was fully randomized and took place on the cut-off date of the treatment. Participants in the treatment group were treated through weekly reports that were sent out at the beginning of each week in a weekly email.[3] The first reports were sent out on 11 October 2021. Unintentionally, each participant's first week's report was sent out again in weeks 2 to 6, instead of providing weekly updated reports. To control for this mistake, robustness checks are conducted in Section 5.5. Hence, the correct treatment period began on 22 November 2021 and lasted until 20 February 2022. The weekly reports were compromised of modular panels, as shown in Figure 2.3. The reports contained various types of information, which aimed at making the email an effective nudge.

Figure 2.3: Example of a weekly report



Firstly, the participant's total external costs of the current week were depicted, separately for each transport mode. The costs were divided into the dimensions health, climate and congestion externalities. Secondly, the participant's total external costs of the current week were

[3]In fact, the emails provided a link, so participants only accessed the reports when clicking on it.

compared to their own external costs of previous study weeks. The baseline consists of one's average external costs between July and September 2021. Throughout the whole treatment period, the baseline was kept fixed and not adjusted if one produced fewer externalities. In addition, the participant's total weekly external costs were compared to other participants' external costs that had recorded the same number of tracking days in the current week, and thus were directly comparable. Based on these two comparisons, each participant received two smileys per week: one for the baseline comparison with one's history of external costs, and one for the comparison with other participants' external costs of the current week. Participants earned a smiley face for doing better than in previous weeks, or for doing better than the comparison group, respectively. Likewise, participants received a frowning face for generating more externalities than previously, or more than the comparable group. Participants who comparatively stayed within a deviation range of max. 10 percent received a neutral smiley. Thus, participants could receive any possible combination between two frowning faces and two smiley faces in total. Lastly, participants were reminded how important reducing external costs was, and that mobility behavior had a strong impact on other people.

3 Methodology

This section explains the methodological concepts behind the results presented in Section 5. In the first part, the identification and estimation of the average treatment effect (ATE) with the Difference-in-Differences (DiD) estimator are explained. This part directly includes the discussion about the validity of the identifying assumptions. The second part explains the willingness-to-pay concept as well as the social welfare analysis that Allcott and Kessler (2019) implement.

3.1 Regression framework

3.1.1 Identifying assumptions

Three assumptions needed to identify treatment effects based on the notation used in Lechner (2011) are presented in the following.

$$\text{SUTVA:} \quad Y_t = dY_t^I + (1-d)Y_t^N \quad \forall t \in \{0,1\} \tag{3.1}$$

The Stable Unit Treatment Value assumption (SUTVA) requires that the treatment of the treated group does not entail spill-over effects on the control group or vice versa. By no means should the control group be affected by the treatment of the treated group. For instance, someone in the control group could benefit from less congested roads due to treated participants not choosing the exact same road at the exact same time anymore. However, the scale of this study makes such scenarios highly implausible, and thus this assumption most likely holds.

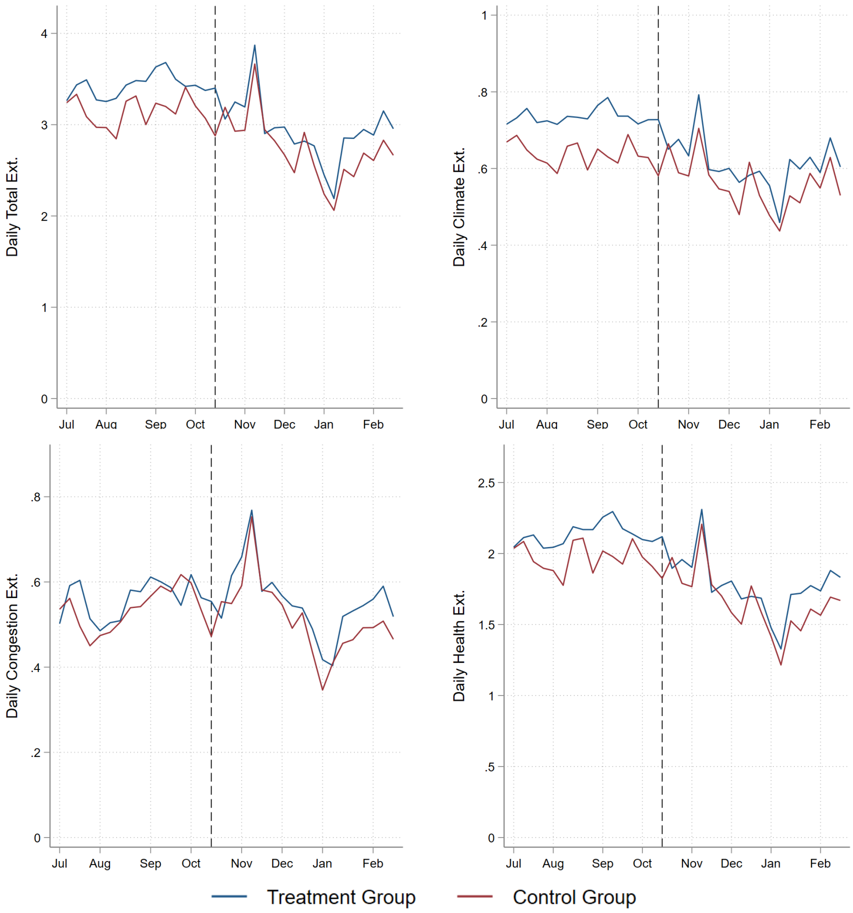
$$\text{No Anticipation:} \quad \mathbb{E}\left[Y_0^I - Y_0^N \mid D=1\right] = \mathbb{E}\left[Y_0^I - Y_0^N \mid D=0\right] = 0 \tag{3.2}$$

The next assumption rules out that individuals anticipate the treatment, which distorts treatment effect estimation by potentially influencing pre-treatment outcomes of both groups. In this study, this assumption is satisfied, as participants did not receive any information about upcoming treatment.

$$\text{Common Trends:} \quad \mathbb{E}\left[Y_1^N - Y_0^N \mid D=0\right] = \mathbb{E}\left[Y_1^N - Y_0^N \mid D=1\right] \tag{3.3}$$

At the basis of the DiD approach lies the assumption that the treated group would follow the same trend as the control group if it had not been treated (Lechner, 2011). To calculate treatment effects, differences between the two groups over time are calculated. This method does not require the treatment and control group to have the same pre-treatment outcomes, but the two groups have to follow a parallel trend. Consequently, any deviation of the trend after the beginning of the treatment between the two groups is directly attributed to treatment. The assumption also rules out that any other policies, characteristics, or effects impact the two groups differently.
Before regressing time trends on external costs, the common trends assumption is analyzed graphically. Figure 3.1 displays average daily external costs per group, averaged over weekly intervals. I choose weekly spans to lessen the noise embodied in individual days. Note that only pre-treatment trends (up to the dashed line) may be compared.

Figure 3.1: Common trends plot



Note: Weekly average externalities per treatment group displayed. Externalities measured in Swiss Francs per day. The dashed line indicates the beginning of the treatment period.

The plots do not correct for date-specific effects, but in expectation, these affect the treated and control group in similar ways. Figure 3.1 indicates that the two groups approximately follow the same trend before treatment begins. The only noteworthy deviation is described by a spike in the treatment group data at the beginning of September in the graphs depicting total, climate and health externalities.

To provide additional evidence, I run regressions on pre-treatment averages, excluding any observations during the treatment period and without fixed effects.[4] Table 3.1 shows that the $Week$ coefficient is significant for one estimate, indicating that there may be a significant time trend. As the interaction between $Treatment$ and $Week$ remains insignificant at the 5%-level for all outcome variables, the trend does not significantly differ between the two groups. Common trend regressions were also conducted on a daily instead of a weekly basis. As these regressions show the exact same insights, they are presented in Table A.1 in the Appendix. Based on the presented analysis, the common trends assumption can be assumed to hold.

[4]Including any of the fixed effects would lead to collinearity.

Table 3.1: Weekly common trends regression, per mode of transport

	(1) Total Ext.	(2) Total Ext.	(3) Total Car Ext.	(4) Total PT Ext.	(5) Total Bike Ext.	(6) Total Walk Ext.
Treatment	0.284+	0.260	0.259	0.003	-0.021	0.005
	(0.171)	(0.202)	(0.209)	(0.033)	(0.018)	(0.013)
Week	-0.001	-0.002	-0.011	0.008***	-0.000	-0.001
	(0.007)	(0.011)	(0.011)	(0.002)	(0.001)	(0.001)
Treatment x Week		0.003	0.009	-0.005+	-0.000	0.000
		(0.014)	(0.014)	(0.003)	(0.001)	(0.001)
adj. R^2	0.00062	0.00061	0.00090	0.00101	0.00086	0.00019
Clusters	777	777	777	777	777	777
N	69219	69219	69219	69219	69219	69219

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). Regressions are run on pre-treatment averages, and only include observations up to the beginning of the treatment period. The dependent variable is aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. No fixed effects are included.

The three assumptions above are enough to estimate the ATE in a DiD setting with randomized treatment. The next section explains how the ATE is estimated in practice in a panel data setting with more than two time periods.

3.1.2 Estimation

The randomized treatment creates an exogenous variation that can be directly used to identify treatment effects (Angrist and Pischke, 2008). Neither endogeneity nor self-selection problems need to be addressed. There is no need for control variables either, as treatment and control groups will be affected equally in expectation (Angrist and Pischke, 2008). The fixed effects panel data estimator used here allows controlling for fixed effects, while comparing average outcomes on a day-specific basis (Cameron and Trivedi, 2005). The estimator implements ordinary least squares regression on the demeaned variables (Cameron and Trivedi, 2005). To estimate treatment effects, data are aggregated to the person-day level. Similar to Axhausen et al. (2021), the ATE can be estimated by

$$Y_{it} = c_0 + \theta^T \cdot DiD_{it}^T + \mu_i + \mu_t + \epsilon_{it} \tag{3.4}$$

The dependent variable is the outcome of interest for person $i \in \{1, \dots, N\}$ on calendar day $t \in \{1, \dots, T\}$. The main dependent variables are the total quantity of external costs, external costs along a particular dimension (health, climate and congestion), and distance traveled (in total or by mode). The DiD term is the product of the treatment group and a treatment period dummy. The term equals one if the treatment (T) is active for a person i on a particular day, and zero otherwise. To control for unobserved heterogeneity, I include fixed effects on person (μ_i) and calendar day (μ_t) level. People who tend to produce more external costs than others regardless of the treatment are assigned a higher value for μ_i. In other words, participants' external costs are compared relative to their external costs. The date fixed effects μ_t capture common shocks that affect mobility behavior throughout the whole country. The error term ϵ has an expected mean of zero and a variance of σ. By including cluster-robust standard errors, I allow for a correlation of the error within participants.
To analyze treatment effect heterogeneity, the DiD term is interacted with dummy variables.

For instance, to investigate potential treatment differences between genders, treatment effects are estimated by

$$Y_{it} = c_0 + \theta^T \cdot DiD_{it}^T + \gamma^T \cdot DiD_{it}^T \cdot male_i + \mu_i + \mu_t + \epsilon_{it} \tag{3.5}$$

where $male_i$ is a dummy variable equal to one for men, and zero otherwise. The ATE for women is thus given by θ^T, whereas the ATE for men is described by $\theta^T + \gamma^T$. The same logic applies to any further interactions.

Even though theoretically, there is no need for control variables, I still include the weather to potentially increase the precision of the estimates. The weather serves as an important predictor of mobility behavior, especially for leisure activities and active transport modes such as bicycling and walking. Tracking data are enriched with temperature and precipitation data from MeteoSwiss.[5] Following Axhausen et al. (2021), I assign weather variables separately for each recorded trip based on a 1 x 1 km grid. To allow for a nonlinear effect of temperature on travel choices (e.g. reflecting the fact that it can be too hot or too cold), $Heat$ and $Cold$ for an observed trip j on day t are defined as follows:

$$Heat_{jt} = \max\left(tmax_{jt} - 25, 0\right) \tag{3.6}$$

$$Cold_{jt} = \max\left(10 - tmind_{jt}, 0\right) \tag{3.7}$$

To compute the values per person and day, averages of the heat, cold and precipitation values across all trips taken by a person i on day t are calculated. Precipitation, heat and cold are then added as linear control variables to (3.4) in some regressions.

Regressions with external costs as the dependent variable are estimated in levels using Stata's reghdfe command. Taking levels (rather than logs) is necessary as external costs associated with walking are negative, and the data provide several person-days with overall negative values (Axhausen et al., 2021). To estimate proportional responses, the coefficients (in Swiss Francs) are divided by the average external costs of the control group generated during the treatment period.

[5] Data are taken from `https://www.meteoswiss.admin.ch`.

3.2 Willingness-to-pay concept

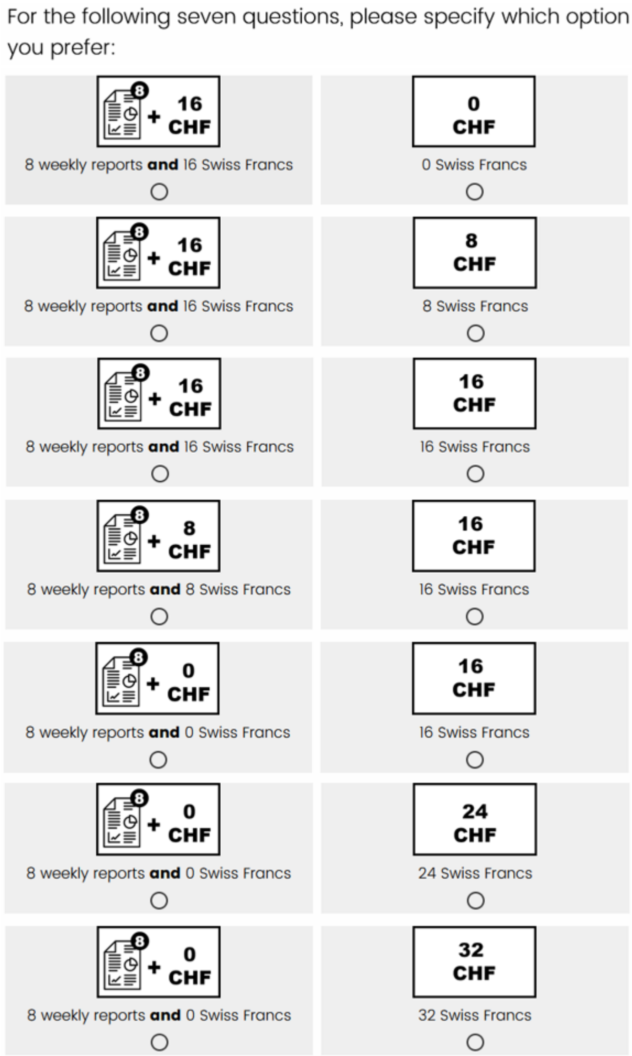
With DiD estimation, the amount of reduced externalities resulting from the intervention can be estimated. However, such an analysis ignores benefits or costs other than mobility externality reductions. For example, nudge recipients may well incur social or psychological costs due to receiving weekly reports (Allcott and Kessler, 2019). Hence, it seems essential to analyze the impact on consumer welfare in order to gain insight into the social welfare consequences of such an intervention. Following the revealed preference approach of Allcott and Kessler (2019), this thesis tries to draw conclusions about consumer welfare by indirectly asking consumers about their WTP. In the next section, their approach is briefly explained.

3.2.1 Eliciting willingness-to-pay

As shown in Figure 3.2, participants are asked to trade off 8 more reports with different amounts of money in the final survey. For instance, participants who prefer "8 weekly reports and 8 Swiss Francs" instead of "16 Swiss Francs" value one additional report at one Swiss Franc or more. Participants who prefer "16 Swiss Francs" instead of "8 weekly reports and 0 Swiss Francs" value one additional report at 2 Swiss Francs or less. A person who answers as in these two examples therefore reveals a WTP between 1 and 2 Swiss Francs per additional report. As some participants opted out of the program even though the reports were free, I allow for revealing negative WTP. For example, participants who choose "8 Swiss Francs" instead of "8 weekly reports and 16 Swiss Francs" are giving up 8 Swiss Francs to *not* receive 8 more reports. Hence, their WTP must be no greater than -1 Swiss Franc per report. Complete and transitive (i.e., internally-consistent) responses to the seven questions in Figure 3.2 allow placing each respondent's WTP into one of the following eight ranges: $(-\infty, -2]$, $[-2, -1]$, $[-1, 0]$, $[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 4]$, and $[4, \infty)$.
To code the WTP from the responses, WTP is assumed to be uniformly distributed on the six interior ranges, and triangularly on the highest and lowest ranges. Assuming triangular distribution is sensible if values lie in the middle of a range with greater probability (Kotz and Van Dorp, 2004). For simplicity, one unique WTP is assigned to each range. For the six interior ranges, the mean of the endpoints is assigned. In other words, all responses on [3, 4] are assigned a WTP of 3.50 Swiss Francs. For the unbounded ranges, i.e., WTP less than -2 Swiss Francs or more than 4 Swiss Francs, the conditional distribution of WTP is assumed to be triangular, with the initial density equal to the average density of the adjoining range. This leads to 12.43 and -3.62 Swiss Francs, respectively, as the conditional mean WTPs on $[4, \infty)$ and $(-\infty, -2]$.[6] The results section also contains estimates under alternative assumptions.
Before calculating mean WTP, it is essential to specify the target population P of the WTP

[6]For example, the density on [3, 4] is 1.67 percent of respondents per Swiss Franc, and the mass in the range $[4, \infty)$ is 21.11 percent of respondents. I assume that these respondents above 4 Swiss Francs are triangularly distributed on $[4, \infty)$, with a maximum density of 1.67 percent per Swiss Franc at 4 Swiss Francs, decreasing to zero density above some upper bound. A simple geometric calculation gives an upper bound of 29.28 Swiss Francs. Following from that, the mean WTP on $[4, \infty)$ is 12.43 Swiss Francs. Analogously, the mean WTP on $(-\infty, -2]$ can be calculated, taking into account that the density on [-2, -1] is 7.78 percent per Swiss Franc, and the mass below -2 Swiss Francs is 18.89 percent. This leads to a lower bound of -6.86 Swiss Francs and a resulting mean WTP of -3.62 Swiss Francs.

Figure 3.2: Seven questions asked in the final survey to elicit WTP



analysis. The overarching goal is to evaluate individual WTP for any one of the treatment group who received weekly reports, regardless of whether or not participants filled out the final survey. Consistently, the WTP analysis should include both treated individuals who opted out of the study and individuals of the treatment group who did not complete the final survey. Similar to Allcott and Kessler (2019), I use inverse probability weights (IPWs) to re-weight the sample to match a target on observable characteristics. This method cannot correct for unobservable characteristics, however. A probit regression presented in Table A.25 in the Appendix estimates

$$Pr(H_i = 1|\boldsymbol{X}_i; P) \tag{3.8}$$

where P denotes the target population, H_i indicates whether person i filled out the WTP questions in the final survey, and $\boldsymbol{X}_i$ includes all observable characteristics used as covariates in the probit regression. The resulting sample weights are described by

$$[\hat{Pr}(H_i = 1|\boldsymbol{X}_i)]^{-1} \tag{3.9}$$

The entire WTP elicitation is incentive-compatible. One of the seven WTP questions was randomly selected for 10 percent of respondents, and they received what they had chosen in that question: either money and/or additional reports.

3.2.2 Welfare analysis

First and foremost, nudging may influence the welfare of study participants. Similar to Allcott and Kessler (2019), this thesis estimates the participant welfare effect ΔV of the mobility externality reports by eliciting WTP for the nudge. This method requires the key assumption that participant i's WTP w_i equals the participant welfare change ΔV_i from the nudging treatment:

$$\Delta V_i = w_i \tag{3.10}$$

Thereby, I assume that the experimental design correctly elicits WTP, and that participants are well-informed and sophisticated in the sense that their WTP for the nudge equals the true effect on their welfare. Such an assumption is only plausible in settings where participants know well about the nudge, and if the nudge brings along behavioral biases, are aware of such biases. These assumptions are discussed in more detail in Section 6.

If a nudge provides valuable information or eliminates bias, it can increase the participants' welfare. Theoretically, the nudge can also decrease the welfare of study participants if it serves as a "moral tax". To estimate the intervention's overall social welfare effect ΔW, other aspects besides ΔV need to be taken into account. ΔW can be denoted as

$$\Delta W = -\Delta x \cdot (p + t - \pi + \phi_e) - C + \Delta V \tag{3.11}$$

where the vector Δx denotes mobility reductions. I quantify benefits from participants' monetary (p is a vector of prices) and time savings (t is a vector denoting saved time), lost profits of gasoline and public transport companies (π denotes a vector of profit margins), and reduced external costs (ϕ_e denotes total external costs saved). The latter is estimated by multiplying the daily average reduction per person by the total amount of treatment days. Long-term externality reductions do not seem to exist and thus do not need to be considered, as section 5.4 will show.[7] The benefit calculations in this thesis neglect private time savings t due to the limited time frame of the thesis. Private monetized cost savings p are partly offset by revenue losses for either gasoline or public transport companies. Yet, lost businesses and forgone mark-ups of gasoline or public transport companies enter the cost side via π.[8] C describes the costs of the intervention, which consist of total implementation costs of the nudge and monetary incentives from the WTP lottery in the final report. In this thesis, I omit the accounting analysis of calculating nudge implementation costs.[9] Therefore, this thesis cannot draw conclusions about the overall welfare effects of the nudging intervention. Nevertheless, an overview and estimation of the sum of consumer welfare change and monetized external cost savings on the benefits side are given in Section 5.6. All benefits are quantified in monetary units (Swiss Francs).

[7] This is in line with the "no persistence" assumption of Allcott and Kessler (2019).

[8] Estimating the lost value added by gasoline and transport companies, which generate less revenue and thus a lower overall mark-up, exceeds the scope of this thesis.

[9] To estimate the cost side correctly, I would need to collect any hourly rates of employees of the ETH Zurich and the University of Basel contributing to this study. This goes beyond the scope of this thesis.

4 Data

4.1 Comparison with Swiss population

Table 4.1 shows that the study sample slightly differs from the general population in certain areas but is otherwise quite representative. The study sample includes a slightly higher share

Table 4.1: Comparison to the Microcensus

		This Study			Microcensus
Category	Level	Control	p-value (Diff.)	Treatment	
Access to car	Yes	76.30	0.547	78.12	75.8
	Sometimes	13.02	0.505	11.45	18.1
	No	10.68	0.912	10.43	6.2
Age	Under 18				13.2
	[18, 25]	8.33	0.231	6.11	9.0
	(25, 35]	13.28	0.933	13.49	14.2
	(35, 45]	23.44	0.193	19.59	15.4
	(45, 55]	21.35	*0.021	28.50	16.7
	(55, 65]	26.04	0.664	24.68	12.9
	66 and older	7.55	0.927	7.63	18.5
Education	Mandatory	3.39	0.394	4.58	19.3
	Secondary	44.79	0.524	47.07	49.5
	Higher	51.82	0.333	48.35	31.2
Employment	Employed	72.66	*0.043	78.88	48.2
	Self-employed	3.39	0.095	1.53	7.2
	Apprentice				2.6
	Unemployed	3.91	0.284	2.54	2.5
	Student	4.95	0.079	2.54	3.7
	Retired	9.38	0.633	8.40	19.3
	Other	5.73	0.824	6.11	16.5
Gender	Female	42.45	0.765	43.51	50.7
	Male	57.55	0.765	56.49	49.3
Household size	1	18.75	0.835	19.34	34.0
	2	37.50	0.906	37.91	35.4
	3	16.15	0.526	14.50	13.0
	4	20.57	0.852	21.12	12.5
	5 or more	6.77	0.707	6.11	5.1
Household income	4'000 CHF or less	5.73	0.286	4.07	17.8
	4'001 - 8'000 CHF	29.17	0.734	30.28	32.8
	8'001 - 12'000 CHF	32.81	0.598	34.61	17.4
	12'001 - 16'000 CHF	14.84	0.453	12.98	6.8
	More than 16'000 CHF	9.38	0.918	9.16	4.5
	Prefer not to say	8.07	0.678	8.91	20.7
Language	German	78.13	0.488	80.15	68.4
	French	18.49	0.418	16.28	25.3
	Italian				6.3
	English	3.39	0.893	3.56	
N		384		393	57'090

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). Except for p-values, all numbers denote percentages.

of (i) older people, (ii) people with higher education, (iii) employed, and (iv) people with higher household incomes. The column between *Control* and *Treatment* tests whether the share of the respective variable significantly differs between the treatment and control groups using a t-test. As the vast majority of comparisons are insignificant, the study sample can be regarded as randomized.

4.2 Data cleaning

Before the analysis, the data were cleaned in order to remove implausible and obviously problematic observations. In particular, I removed the data if one of the following things occurred:

- Average daily speed above 100 km/h over all modes of transport
- Average daily speed above 100 km/h for car and public transport, above 40 km/h for bicycling, and above 20 km/h for walking
- More than 500 km/day over all modes of transport
- More than 500 km/day for car and public transport, more than 150 km/day for bicycling, and more than 20 km/day for walking
- Total duration of travels above 10 hours on a specific day

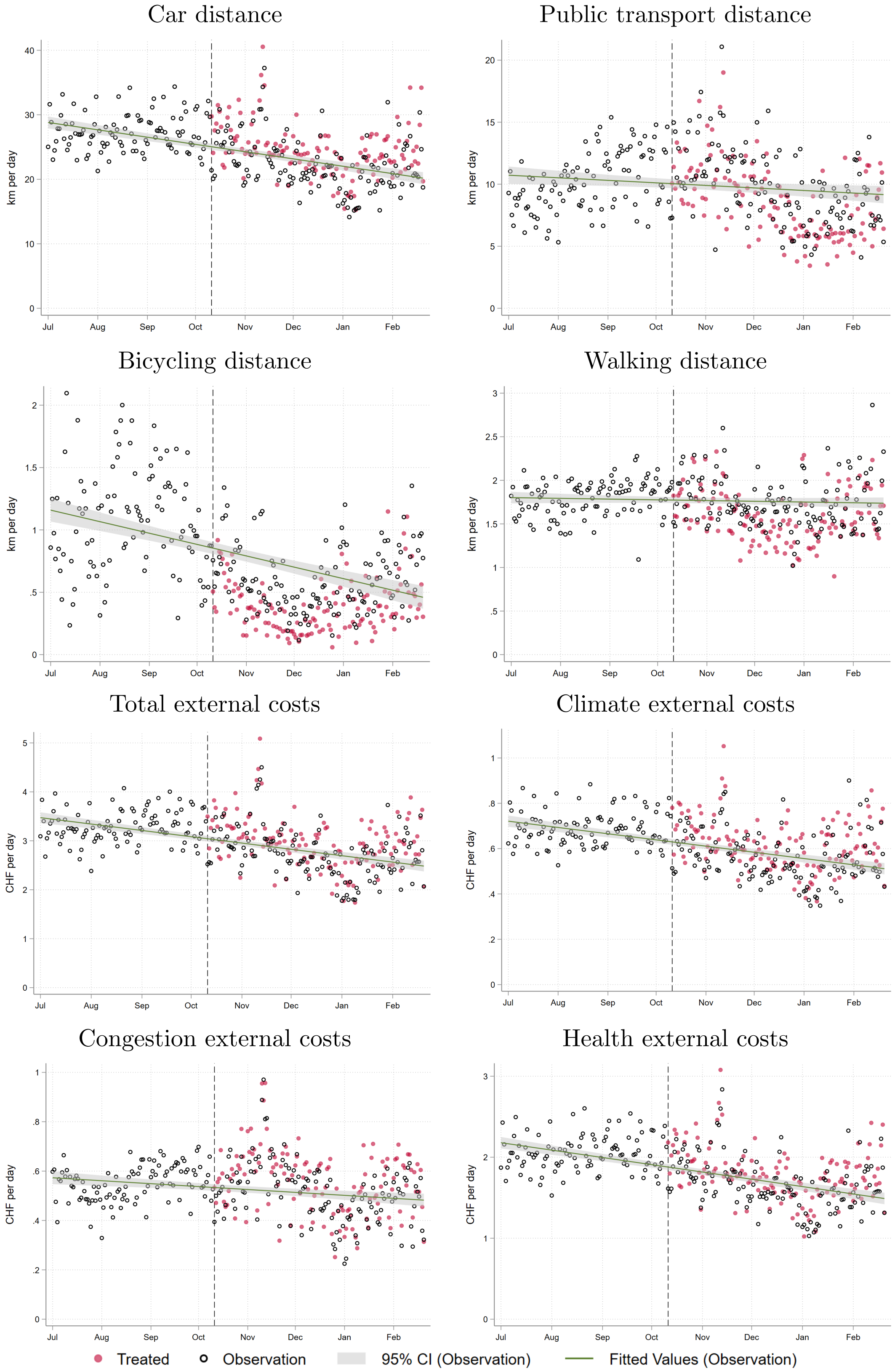
To not cause any distortion, I removed all data for that person and that day from the analysis.

Moreover, all participants with less than four tracking days over the entire treatment period were excluded from the regressions due to lack of variation.

4.3 Tracking statistics

Figure 4.1 presents tracking statistics over the course of the experiment. The figures plot average values over treatment status and calendar date. The "observation" dots include observations from the control group plus observations from the treatment group before the treatment. The plots display a fitted line over all observation group values, enclosed by 95%-confidence intervals. All plots in Figure 4.1 depict a negative time trend, indicating that there is seasonal variation in travel distance by mode. The negative trend is particularly strong for bicycling. The seasonal variation in travel distance translates to a negative seasonal trend in external costs. Without a control group, the negative trend could be wrongly interpreted as treatment effect. The outliers in Figure 4.1 also highlight the need for a control group, which is exposed to the same unobserved shocks as the treatment group. Including calendar date fixed effects controls for unobserved characteristics that give rise to the time trend in the data, but which are unrelated to the treatment.

Figure 4.1: Tracking statistics



5 Results

5.1 Descriptive statistics

Table 5.1 shows tracking summary statistics for overall travel. Average values per participant and day are calculated for the pre-treatment as well as the treatment period.

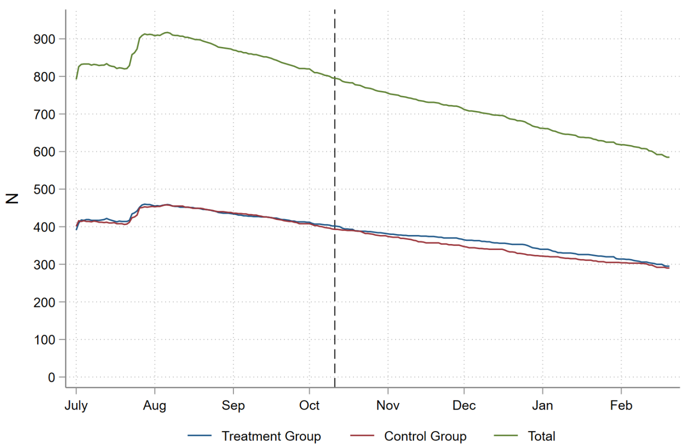
Table 5.1: Summary statistics

		Pre-treatment		Treatment	
Dimension	Outcome	Control	Treatment	Control	Treatment
Ext. costs	Total	3.348	3.631	3.012	3.278
(CHF)		(5.434)	(5.702)	(5.097)	(5.368)
	Congestion	0.573	0.593	0.563	0.613)
		(1.236)	(1.185)	(1.128)	(1.202)
	Climate	0.679	0.778	0.616	.683
		(1.235)	(1.353)	(1.216)	(1.282)
	Health	2.096	2.260	1.833	1.982
		(3.574)	(3.727)	(3.336)	(3.437)
Tracking	Distance (km)	43.077	44.553	38.483	39.469
		(60.052)	(61.618)	(55.794)	(58.001)
	Duration (min)	85.174	86.025	75.772	74.132
		(82.901)	(83.220)	(75.063)	(73.511)

Notes: Average values per participant and day during the experiment. Standard errors in parentheses.

Figure 5.1 displays the total amount of participants in both groups over time. At the beginning of the control period, 826 participants formed part of the experiment. Due to the ongoing recruitment at the beginning (and some opting out already), this number rose to 917 participants at the beginning of August 2021. Participants continuously opted out of the experiment in the following months. At the end of the treatment period, a total of 585 participants were still tracking, divided equally into treatment and control groups. Importantly, there was no difference in attrition. Following the approach by Macours and Molina Millán (2017), I use probit regressions to estimate the propensity to continue the study. Table A.2 in the Appendix shows that the coefficient *Treatment Group* is not significant.

Figure 5.1: Amount of study participants over time



5.2 Average treatment effects

Table 5.2 shows the average treatment effect (ATE) on the external costs of mobility in Swiss Francs per day. Due to the randomization and the presence of the control group, the treatment coefficients can be interpreted as the causal effects of nudging.

Table 5.2: Average treatment effects on external costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Ext. Cost		Climate Ext. Cost		Congestion Ext. Cost		Health Ext. Cost	
Treated	-0.058	-0.060	-0.035*	-0.036*	0.015	0.016	-0.038	-0.039
	(0.084)	(0.084)	(0.020)	(0.020)	(0.022)	(0.022)	(0.053)	(0.053)
Heat		-0.377***		-0.102***		-0.005		-0.270***
		(0.054)		(0.013)		(0.009)		(0.037)
Cold		0.087***		0.035***		-0.025***		0.077***
		(0.020)		(0.005)		(0.003)		(0.014)
Precipitation		-0.002		0.000		-0.002**		0.000
		(0.004)		(0.001)		(0.001)		(0.003)
adj. R^2	0.176	0.177	0.168	0.170	0.150	0.151	0.181	0.182
Clusters	777	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on one-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. Precipitation is measured in mm per hour; heat and cold are defined as in (3.6) and (3.7). All regressions include date and person fixed effects.

The first two columns report the results for the total external costs of transport, with and without controlling for the weather. Columns 3 to 8 show the ATE on climate, congestion and health externalities. Column 2 estimates that nudging reduces total externalities by 6 cents per day on average, but not significantly at conventional significance levels. Column 4 states that treatment significantly reduces daily climate externalities by 3.5 cents on average ($p < 0.05$ with one-sided testing). Column 6 estimates that nudging does not affect congestion externalities in any way. The effect magnitude is weak and close to zero. Finally, column 8 shows that the treatment reduces health externalities, but again not to a significant extent. Including the weather barely alters the ATE.[10]

[10] All estimates in Section 5.3 exclude the weather, as including it hardly affects the coefficient for *Treated*.

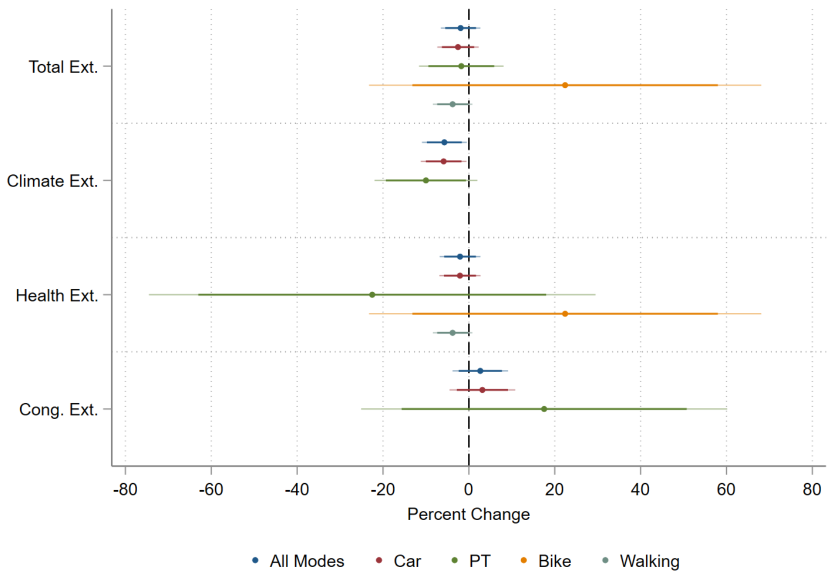
Table 5.3: ATE: Proportional responses

	(1) Total Ext.	(2) Climate Ext.	(3) Total Car Ext.	(4) Climate Car Ext.	(5) Total PT Ext.	(6) Climate PT Ext.
Treated	-0.019	-0.057*	-0.025	-0.059*	-0.018	-0.100+
	(0.028)	(0.032)	(0.029)	(0.032)	(0.060)	(0.073)
adj. R^2	0.176	0.168	0.187	0.174	0.163	0.108
Clusters	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on one-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. To estimate proportional responses, coefficients are divided by the average external costs of the control group during the treatment period. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Table 5.3 and Figure 5.2 both display the proportional reduction of the external costs of transport, with bars for 95%-confidence intervals. The respondents in the treatment group reduce their climate externalities by 5.7% relative to the control group. This is one of the core results of the analysis and statistically significant at $p < 0.05$. Treatment reduces total external costs by 1.9%, but not significantly at usual levels. Figure 5.2 presents large confidence intervals for bike externalities. This variability occurs because many study participants hardly ever ride a bike.

Figure 5.2: Proportional treatment effects, by mode



5.3 Effect heterogeneity

Estimating ATE provides an important initial overview of the overall effect. However, estimating total effects may conceal the fact that certain subgroups respond differently to the treatment. Heterogeneous treatment effects could then provide valuable further insights. This section presents several heterogeneity analyses: First, it shows treatment effects for those subgroups of the sample also analyzed in Axhausen et al. (2021). Second, it analyzes trends of the treatment effect over time. Third, it examines the static and dynamic psychological impact of previously received nudging smileys.

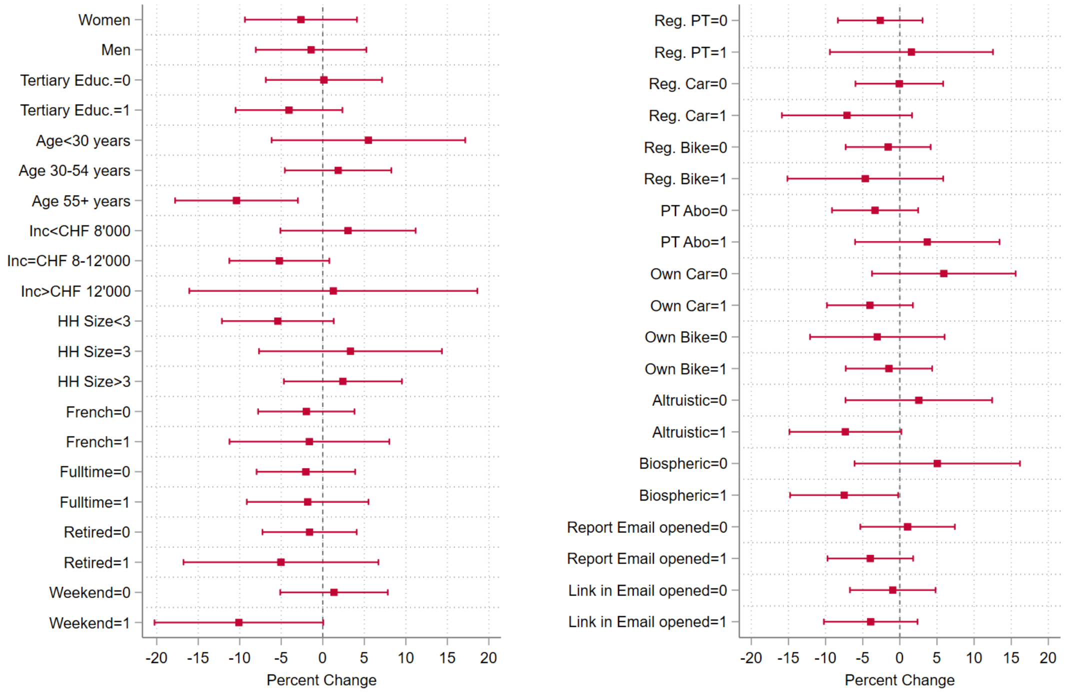
Figure 5.3 presents total relative treatment effects on total and climate external costs for different subgroups of the sample.[11] All corresponding regression results as well as Figures containing only the interaction terms are shown in the Appendix. Figure 5.3 illustrates that (i) people aged 55 years or above and (ii) participants on weekends respond significantly to the treatment. The final survey after the experiment contained a battery of questions to derive respondents' personal values (Schwartz, 1992; Steg and De Groot, 2010). Using this methodology, participants were assigned an index along the dimensions "altruistic" and "biospheric".[12] People that scored above the median in terms of the biospheric index significantly reduce their external costs. Likewise, respondents classified as "altruistic" responded significantly to the treatment. For socio-demographic variables not mentioned so far, I do not find statistically significant differences. There is no significant effect heterogeneity in terms of gender, education, household size, and language regions.
In contrast to total external costs, (i) people living in a small household with a maximum of two persons, (ii) people that regularly use a car or (iiii) own a car, and (iv) people classified as altruists significantly reduce their external costs. Moreover, (v) people aged 55 years or above, (vi) people earning between 8'000 and 12'000 CHF per month, (vii) people on weekends, and (viii) people that do not own a public transport subscription react significantly to the treatment. Moreover, participants who had opened the report email in the previous week significantly reduce their externalities in the week after having opened the email. Similarly, a significant reduction results for people having opened the link in the report email, with which participants directly access their personalized report. This interaction allows distinguishing the effect between people that were solely reminded of the "pill" via the email, and people that actually took the "pill" when opening the link.
Table 5.4 presents multivariate regression estimates, which include 13 potentially relevant categorical variables that are interacted with the DiD term. Which variables to include was decided based on economic theory and the results of the univariate interaction regressions displayed in Figure 5.3. In the multivariate regressions, all dimensions include one omitted category. The "base" coefficient is thus associated with an observation that has a zero for all included dummies.

[11] To estimate the correct treatment effects in percentages, I would theoretically have to divide absolute reductions by the average external costs of the respective average of the control-subgroup during the treatment period (instead of dividing by the average over all participants of the control group). This small divergence is neglected here, as this is not the main focus of this thesis.

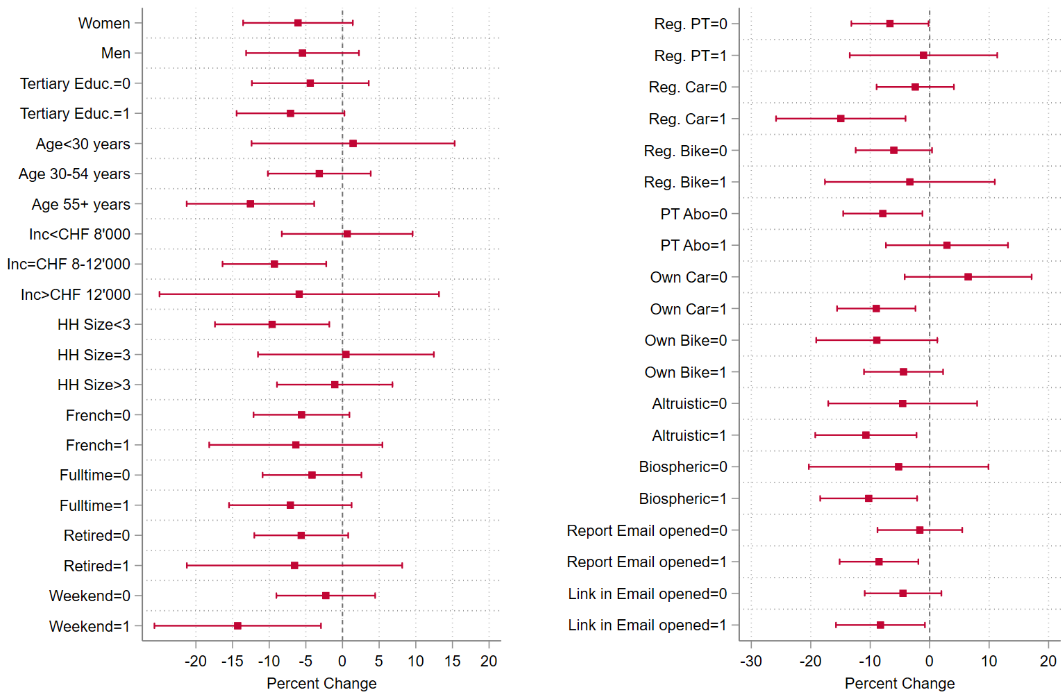
[12] Instead of eliciting all 4 dimensions, only 8 questions for the two values "altruistic" and "biospheric" were asked. However, a series of robustness checks were conducted to make sure that this tightening acts in no way distorting.

Figure 5.3: Effect heterogeneity: Proportional total responses

Total external costs



Climate external costs



Note: Total heterogeneous treatment effects (in percent) on total and climate external costs are shown. 95%-confidence intervals are displayed.

Table 5.4: Multivariate interactions

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.247	0.032	0.074	0.141
	(0.203)	(0.044)	(0.050)	(0.129)
Male=1	0.133	0.030	0.022	0.080
	(0.138)	(0.031)	(0.041)	(0.086)
Tertiary educ.=1	-0.110	-0.009	-0.066*	-0.035
	(0.114)	(0.027)	(0.030)	(0.072)
Age<30	0.109	0.030	-0.045	0.125
	(0.181)	(0.044)	(0.044)	(0.111)
Age>55	-0.401**	-0.061+	-0.089*	-0.251**
	(0.146)	(0.034)	(0.037)	(0.096)
HH Inc.<8'000 CHF	0.283*	0.070*	0.042	0.171*
	(0.133)	(0.031)	(0.030)	(0.086)
HH Inc.>12'000 CHF	0.241	0.025	0.074	0.142
	(0.270)	(0.060)	(0.084)	(0.153)
HH size<3	-0.236	-0.060	-0.045	-0.131
	(0.179)	(0.041)	(0.045)	(0.110)
HH size>=4	-0.056	-0.008	-0.054	0.006
	(0.184)	(0.042)	(0.046)	(0.115)
French=1	0.063	0.006	0.007	0.051
	(0.143)	(0.037)	(0.031)	(0.092)
Fulltime=1	-0.035	-0.023	0.062	-0.074
	(0.147)	(0.033)	(0.039)	(0.093)
Retired=1	0.248	0.052	0.007	0.189
	(0.203)	(0.051)	(0.050)	(0.124)
Weekend=1	-0.341+	-0.073+	-0.033	-0.236+
	(0.184)	(0.038)	(0.039)	(0.122)
Email opened=1	-0.142+	-0.042*	-0.006	-0.094+
	(0.076)	(0.018)	(0.018)	(0.048)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. All dimensions include one omitted category. *Treated* is thus associated with an observation that has a zero for all included dummies. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

When controlling for several dimensions simultaneously, column 1 shows that the nudging effect on total external costs is stronger (i.e., more negative) for (i) participants aged 55 years or above, (ii) on weekends, and (iii) for people who had opened the email containing the report, and it is weaker for (iv) participants with a household income below 8'000 Swiss Francs per month. Column 2 shows that the effect on climate externalities is stronger for (i) participants aged 55 years or above, (ii) on weekends, and (iii) for participants having opened the email with the report, and it is weaker for (v) participants with a household income below 8'000 Swiss Francs per month.

As a first approach to analyze the development of treatment effects, an event study design is chosen. However, estimating an individual treatment effect for each study week leads to very noisy estimates due to lack of statistical power. Therefore, these graphs are displayed in Figure A.2 in the Appendix.
Table 5.5 includes linear and quadratic terms for each study week. As participants may become used to receiving weekly reports, treatment effects might flatten over time (Bonezzi et al., 2011). Columns 1 to 8 include a linear study week trend, whereas columns 5 to 8 additionally include a quadratic study week term.

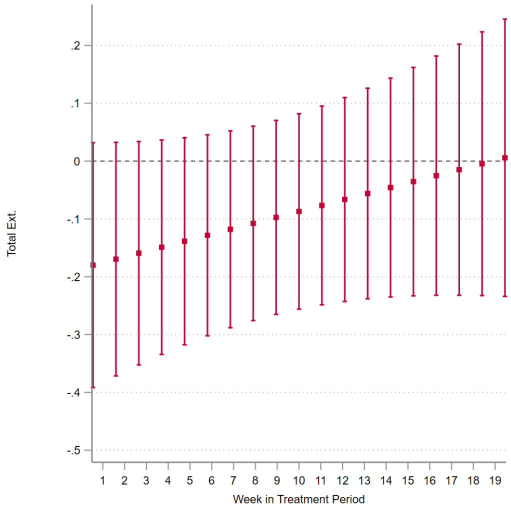
Table 5.5: ATE with linear and quadratic study week terms

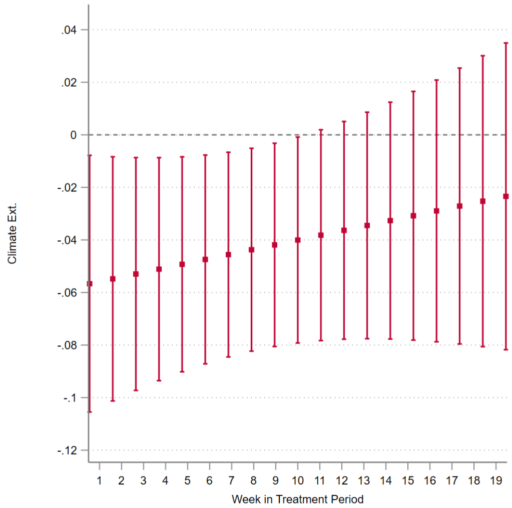
	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.	(5) Total Ext.	(6) Climate Ext.	(7) Cong. Ext.	(8) Health Ext.
Treated	-0.379+	-0.093+	-0.003	-0.283*	0.242	0.086	0.031	0.124
	(0.222)	(0.054)	(0.053)	(0.138)	(0.949)	(0.227)	(0.210)	(0.619)
Treated x Week	0.013	0.002	0.001	0.010+	-0.039	-0.013	-0.002	-0.024
	(0.009)	(0.002)	(0.002)	(0.005)	(0.078)	(0.019)	(0.017)	(0.051)
Treated x Week^2					0.001	0.000	0.000	0.001
					(0.002)	(0.000)	(0.000)	(0.001)
adj. R^2	0.176	0.168	0.150	0.181	0.176	0.168	0.150	0.181
Clusters	777	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Figure 5.4 presents the total treatment effect for each week of the treatment period separately, after including linear study week terms. The total effect on climate externalities is significant and negative for the first 10 weeks, with a clear upward trend.

Figure 5.4: Weekly ATE with linear study week terms

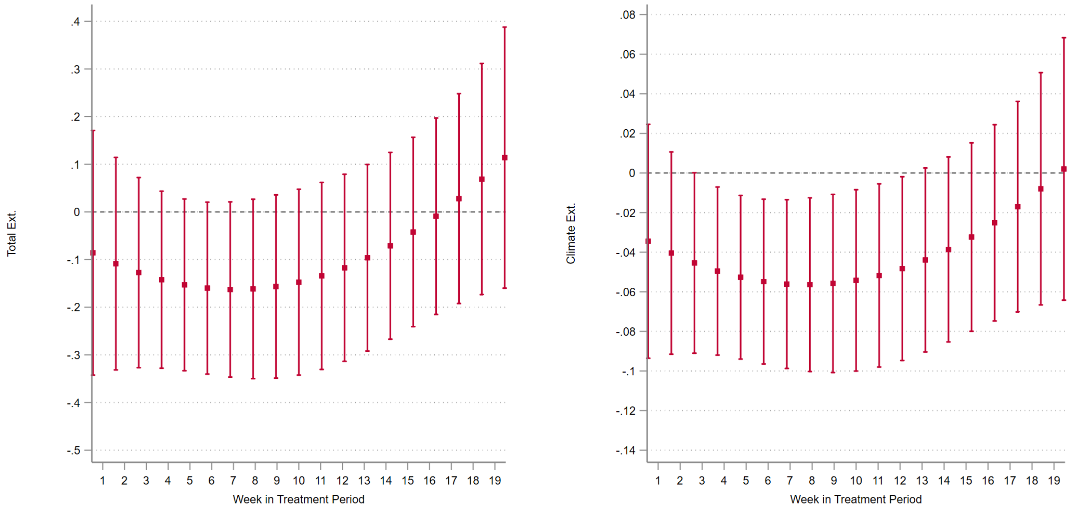




Note: Total treatment effect per week displayed, including a linear term for study week. 95%-confidence intervals are shown.

Figure 5.5 displays the total treatment effect per week when including linear and quadratic study week terms. When comparing the left- and right-hand side, two points stand out: First, the absolute overall effect is much larger for total externalities on the left-hand side. Second, the overall effect on climate externalities is significant within the period from week 3 to week 12, but not on total externalities. The u-shaped trend visible in both graphs is a strong indication that the treatment effect weakens over time.

Figure 5.5: Weekly ATE with linear and quadratic study week terms



Note: Total treatment effect per week displayed, including a linear and a quadratic term for study week. 95%-confidence intervals are shown.

The flattening effect may not only be seen over the whole study period but even within a week. According to Bonezzi et al. (2011), motivation to engage in goal-consistent behavior can be higher when people are either far from or close to the end state. To analyze whether people are most motivated directly after having received the weekly reports (which also serve as a reminder), I run regressions with interactions between each weekday (Tuesday to Sunday, and Mondays being the reference category), and the DiD term. However, as I do not find any noteworthy effect in the data, the regression estimates can be found in the Appendix.

The existing literature in the field of psychology emphasizes that the type of feedback given strongly influences how motivated people are, for example to reduce their externalities further (Belschak and Den Hartog, 2009; Fishbach et al., 2010). Hence, the either happy or frowning smileys in last week's report may influence mobility behavior throughout the entire week (Fishbach et al., 2010). Table 5.6 examines whether the treatment effects differ with respect to what kind of smileys participants received in last week's report. Columns 3, 5 and 6 solely include the baseline comparison. This comparison might be more essential, as many participants may have fewer feelings of inequality and thus more motivation to improve themselves (Rela, 2022). Besides, someone who depends on regular job-related driving does not perform well in the social comparison but still has the opportunity to challenge oneself. The interaction terms include all observation days after sending the report up to

Table 5.6: Average treatment effects with static smiley effects

	(1) Total Ext.	(2) Total Ext.	(3) Total Ext.	(4) Total Ext.	(5) Total Ext.	(6) Total Ext.	(7) Total Ext.
Treated	-0.058	0.167+	0.196*	-0.215**	-0.224**	-0.018	-0.036
	(0.084)	(0.088)	(0.089)	(0.083)	(0.083)	(0.087)	(0.086)
Treated x ☺(☺)		-1.072***					
		(0.079)					
Treated x ☺ (Base)			-1.102***			-0.802***	
			(0.076)			(0.074)	
Treated x ☹(☹)				2.721***			
				(0.151)			
Treated x ☹ (Base)					2.180***	1.897***	
					(0.138)	(0.138)	
Treated x ☺☺							-1.193***
							(0.083)
Treated x ☺							-0.260*
							(0.112)
Treated x ☹							1.111**
							(0.388)
Treated x ☹☹							2.775***
							(0.160)
adj. R^2	0.176	0.178	0.178	0.180	0.179	0.180	0.181
Clusters	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

the date of next week's report. All estimates are regressed on total external costs.[13] The coefficients in Table 5.6 can be interpreted as causal treatment effects. To obtain the total treatment effect of people having received a certain smiley, the treatment coefficients still need to be added up. Column 1 re-displays the baseline estimate from Table 5.2, without any new interactions. Column 2 adds an interaction between the DiD term and a dummy for people having received a happy smiley in their report. More precisely, the dummy includes any combination of neutral and happy smileys entailing at least one happy smiley. After including this variable, the baseline treatment coefficient changes sign. Column 2 estimates that reports with at least one happy smiley lead to participants reducing externalities by 1.07 Swiss Francs per day on average in the following week. Column 3 estimates that people who receive a happy smiley in the baseline comparison to their history of external costs, significantly reduce externalities by 1.10 Swiss Francs relative to the reference group. Column 4 shows that receiving at least one frowning smiley leads to significantly higher external costs in the following week. The magnitude of the increase is extensive, with an average externality increase of 2.72 Swiss Francs per day. Interestingly, *Treated* becomes not only larger but also highly significant in columns 4 and 5. Finally, column 7 combines four possible smiley versions. All coefficients are highly significant, with signs as in the estimates before.

[13]Estimates with climate external costs as the dependent variable can be found in the Appendix.

Table 5.7 looks at dynamic effects within the type of smileys that people receive in the reports. Fishbach et al. (2010) state that shifting from positive to negative feedback or vice versa significantly influences participants' motivation. For this purpose, the smileys in the report of the week before are used to calculate the absolute change in smileys from one week to the next. Column 2 shows that people who receive two happy smileys after having received two frowning smileys further reduce their external costs by 1.55 Swiss Francs on average in the following week. Like all interactions with dynamically changing smileys in Table 5.7, this result is highly significant. Column 3 states that participants who receive one happy smiley after having received one frowning smiley significantly reduce their externalities in the following week, but to a lesser extent than people who moved from two frowning to two happy smileys. As shown in column 4, participants who first receive a frowning and then a happy smiley in the baseline comparison also significantly reduce external costs in the week after. On the contrary, columns 5 and 6 show that participants who receive a frowning after a happy smiley strongly increase their external costs as a response to the smiley degradation. Column 7 shows that people who alter from a happy to a frowning smiley when being compared to their historical costs, significantly increase external costs by 2.03 Swiss Francs in response to the smiley change.

Table 5.7: Average treatment effects with dynamic smiley effects

	(1) Total Ext.	(2) Total Ext.	(3) Total Ext.	(4) Total Ext.	(5) Total Ext.	(6) Total Ext.	(7) Total Ext.
Treated	-0.058	-0.040	-0.037	-0.033	-0.085	-0.140+	-0.154+
	(0.084)	(0.085)	(0.085)	(0.085)	(0.084)	(0.084)	(0.085)
Treated x (☹☹ to ☺☺)		-1.546***					
		(0.159)					
Treated x (☹(☹) to ☺(☺))			-0.713***				
			(0.162)				
Treated x (☹ to ☺ (Base))				-0.535***			
				(0.127)			
Treated x (☺☺ to ☹☹)					2.439***		
					(0.204)		
Treated x (☺(☺) to ☹(☹))						2.615***	
						(0.162)	
Treated x (☺ to ☹ (Base))							2.025***
							(0.147)
adj. R^2	0.176	0.176	0.176	0.176	0.177	0.178	0.178
Clusters	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

5.4 Long-term treatment effects

This study provides a unique opportunity to examine long-term treatment effects, as tracking is continued after the end of the treatment period. The post-treatment period begins on 21 February 2022. For this thesis, observations up to 31 May 2022 are considered, even though tracking still goes on at the time of writing.

Figure 5.6 shows the steadily declining total amount of participants over the pre-treatment, treatment and post-treatment period. Participants keep leaving the study during the post-treatment period, and thus the total amount of participants is further declining. Table A.2 in the Appendix shows that the attrition does not significantly differ between the treatment and the control groups. Post-treatment tracking statistics are presented in Figure A.4 in the Appendix.

Figure 5.6: Amount of study participants over time

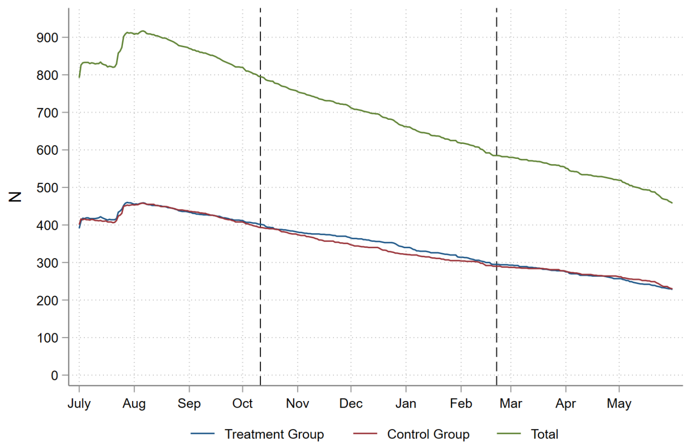


Table 5.8 shows regression results of the post-treatment analysis, including *Treatment* and *Post-treatment* dummies for observations during and after the treatment period, respectively.

Table 5.8: Post-treatment regressions

	(1) Total Ext.	(2) Clim. Ext.	(3) Cong. Ext.	(4) Health Ext.	(5) Total Car Ext.	(6) Clim. Car Ext.
Treated x Treatment	-0.058	-0.035*	0.015	-0.038	-0.072	-0.034*
	(0.084)	(0.020)	(0.022)	(0.053)	(0.085)	(0.019)
Treated x Post-Treatment	0.052	-0.004	0.003	0.052	0.017	-0.006
	(0.102)	(0.025)	(0.023)	(0.065)	(0.100)	(0.024)
adj. R^2	0.184	0.175	0.154	0.188	0.196	0.184
Clusters	777	777	777	777	777	777
N	190284	190284	190284	190284	190284	190284

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on one-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

The point estimates for treatment effects in the first row only marginally change when including the interaction between the DiD term and the post-treatment dummy. According to the second row in Table 5.8, the post-treatment coefficients of all six columns are not statistically significant. Hence, the post-treatment externalities do not statistically differ from the pre-treatment level, which serves as the reference category. However, not seeing and measuring a significant effect in the data does not necessarily mean that there is no long-term effect.

5.5 Robustness tests

This section provides three kinds of robustness checks to assess the main results even better. First, the wrong treatment during the first six weeks is addressed. Next, I discuss the conflict regarding participants having already been part of the previous MOBIS study. Finally, I look more closely at how to treat zeroes in the data.

An important part of the robustness analysis relates to the incorrect reports that were sent out in the first six weeks (see also Section 2.4). Table 5.9 includes two dummies - one for the first six weeks of the treatment period, and one for the remaining weeks of the treatment period. Consequently, the interaction between the dummy *Week 1-6* and the DiD term estimates the average treatment effect over the first six weeks. Likewise, the interaction with *Week 7+* estimates the average treatment effect over the treatment period when excluding the first six weeks. Column 1 in Table 5.9 shows that the magnitude of the effect is much larger within the first six weeks. The effect from week 7 onward still comes with a negative sign, but the effect size is weaker and the p-value much larger. In column 2, the effect on climate external costs is significant during the first six weeks only. This means that the main part of the overall effect on climate externalities can be attributed to the first few weeks. Column 4 shows that the treatment effect on health external costs is both much stronger in effect size and significant at the 10%-level within the first six weeks. Columns 1 to 4 in Table 5.9 allow drawing two conclusions. First, further evidence of a treatment effect that diminishes over the treatment period is provided. Second, the hypothesis is supported that receiving a report is already of great importance, regardless of whether it is read or not. For these two reasons, the first few weeks are included in all further analyses.

Table 5.9: Robustness regressions

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.	(5) Total Ext.	(6) Climate Ext.	(7) Total Ext.	(8) Climate Ext.
Treated x Week 1-6	-0.140	-0.048*	0.010	-0.102				
	(0.103)	(0.024)	(0.026)	(0.064)				
Treated x Week 7+	-0.007	-0.027	0.018	0.002				
	(0.093)	(0.022)	(0.025)	(0.059)				
Treated					-0.092	-0.041+	-0.116	-0.045*
					(0.102)	(0.025)	(0.092)	(0.022)
Treated x MOBIS					0.076	0.012		
					(0.120)	(0.028)		
Treated x MOBIS Info.							0.208	0.037
							(0.134)	(0.030)
adj. R^2	0.176	0.168	0.150	0.181	0.176	0.168	0.176	0.168
Clusters	777	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Another inevitable robustness check is needed because some of the participants had already been part of the MOBIS study by Axhausen et al. (2021). Their RCT also deals with reducing mobility externalities after having received an information (or pricing) treatment. Since the treatment was completely randomly assigned in both experiments, it is theoretically possible that a participant was assigned to the treatment group twice. Therefore, I conduct two kinds of interactions: First, regressions with an interaction between the DiD term and a dummy for all former MOBIS participants are run. Second, I conduct regressions with an interaction dummy for former participants of any of the two treatment groups only. Columns 5 to 8 in Table 5.9 clearly state that none of the interactions with former MOBIS participants is anywhere near significant. Thus, the MOBIS treatment does not seem to have any impact.

Finally, the robustness section concludes with a series of intensive regressions that deal with the treatment of missing tracking data. The tracking app sometimes has troubles distinguishing between missing tracking data and actual zeroes. There is no way of certainly detecting whether the tracking app failed to work properly, whether participants on purpose turned off the tracking app, or whether they really did not move. Since a value of zero is transmitted in all these cases, inexact measurements could distort the results in one or another direction. A couple of regressions using only data from days with nonzero travel distances were run to check for this. Table 5.10 shows that the treatment effects on total and climate externalities are still negative, but become smaller in absolute terms. Besides this logical consequence of excluding zeroes, the results are overall quite stable.
The treatment effect on total externalities remains insignificant, whereas the effect on climate externalities is now significant at the 10%-level. The static and dynamic smiley regressions of Section 5.3 are also run as intensive regressions. These regressions show almost identical point estimates and are therefore not depicted in this thesis. To conclude, the intensive regressions show no serious change in results.

Table 5.10: Intensive regressions

	(1) Total Ext.	(2) Total Ext.	(3) Climate Ext.	(4) Climate Ext.	(5) Cong. Ext.	(6) Cong. Ext.	(7) Health Ext.	(8) Health Ext.
Treated	-0.008	-0.010	-0.027+	-0.027+	0.018	0.019	-0.001	-0.038
	(0.085)	(0.085)	(0.020)	(0.020)	(0.024)	(0.024)	(0.055)	(0.053)
Heat		-0.315***		-0.095***		0.018*	-0.238***	
		(0.062)		(0.014)		(0.010)	(0.043)	
Cold		0.096***		0.039***		-0.029***	0.086***	
		(0.024)		(0.006)		(0.004)	(0.016)	
Precipitation		0.002		0.001		-0.002**	0.002	
		(0.005)		(0.001)		(0.001)	(0.003)	
adj. R^2	0.195	0.196	0.187	0.189	0.166	0.167	0.202	0.181
Clusters	777	777	777	777	777	777	777	777
N	124857	124857	124857	124857	124857	124857	124857	145569

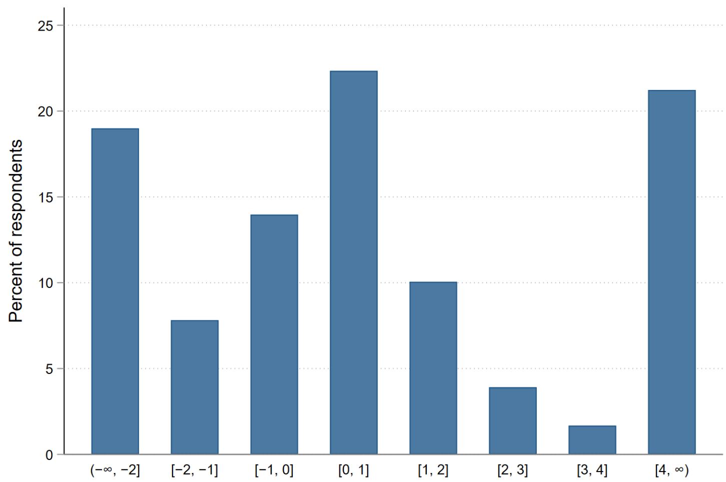
Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on one-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

5.6 Willingness-to-pay estimation

Responses in the final survey that are considered for the WTP estimation need to be both complete and internally consistent. All received responses are complete, whereas only 76.5 percent of the responses are consistent. The 23.5 percent of responses that are complete but internally inconsistent are not further considered in the entire WTP analysis.

Figure 5.7 presents the distribution of WTP, with all responses weighted equally. Respondents are assigned to 8 different ranges, as explained in Section 3.2.1. Across all respondents, 59.22 percent reported a positive WTP. 35.8 percent of all positive WTPs are situated in the highest range, having a minimum WTP of 4 Swiss Francs per report. However, 40.8 percent of all respondents reported a negative WTP. 46.6 percent of all negative WTPs are located in the lowest range, with a resulting WTP of -2 Swiss Francs per report, or less.

Figure 5.7: WTP for additional reports



Notes: The histogram presents WTP per additional report for survey respondents, with all responses weighted equally.

Table 5.11 presents estimates of mean WTP, with standard errors in parentheses. Column 1 presents unweighted estimates, while column 2 uses IPW to match the target population of all treated participants that received at least one report. Re-weighting is necessary, as not all participants of the treatment group filled in the final survey, and thus did not answer questions about their WTP. The probit estimates used for calculating weights are displayed in Table A.25 in the Appendix. As Table 5.11 shows, all estimates entail a positive mean WTP. The unweighted base case estimates a mean of 2.18 Swiss Francs per additional report. When re-weighted on observables to match the target population, the mean rises to 2.42 Swiss Francs per report. The fact that weighting slightly rises mean WTP suggests that final survey respondents are slightly negatively selected on observables.
The seven questions in the final survey allow bounding each respondent's WTP, but the base case estimate makes particular assumptions to go from bounds to point estimates. The next five rows of Table 5.11 consider sensitivity to alternative assumptions.

Table 5.11: Mean WTP estimates

	(1) Unweighted	(2) Weighted
Base case (Triangular distribution)	2.183	2.421
	(0.420)	(0.718)
Uniform WTP at endpoints	1.813	1.979
	(0.353)	(0.606)
WTP = {-5, 12} at endpoints	1.830	2.053
	(0.430)	(0.724)
WTP = {-5, 10} at endpoints	1.405	1.560
	(0.374)	(0.627)
WTP = {-4, 10} at endpoints	1.595	1.750
	(0.357)	(0.605)
WTP bounds closest to 0	0.620	0.647
	(0.155)	(0.259)
Nonrespondents have WTP = 0	0.955	
	(0.191)	

Notes: Standard errors in parentheses.

Rows 2 to 5 in Table 5.11 implement alternative assumptions for mean WTP at the outer two bounds, i.e. mean WTP for respondents with WTP below -2 or above 4 Swiss Francs. The second row assumes uniform distribution at the endpoints, with the density equal to the density of the adjoining range. Mean WTPs are then 10.32 and -3.21 Swiss Francs for the upper and lower bounds, respectively. The next three rows use heuristic benchmarks. Similar to Allcott and Kessler (2019), I choose benchmarks that are located in a similar range as the means of the outer intervals when assuming triangular distribution. Since the adjacent region of the lowest interval (-∞, -2], has a greater density than the adjacent region of the highest interval, the lower bound is reached more quickly. Therefore, the negative lower bound is closer to the adjoining range than the positive upper bound in absolute values. Overall, Table 5.11 indicates that the estimated mean WTP is highly sensitive to assumptions about the two endpoint intervals. This makes intuitive sense, as 40.2 percent of respondents' WTP are located at one of the endpoints. Weighted results vary between 1.56 and 2.42 Swiss Francs when making different assumptions. Unweighted results vary between 1.41 and 2.18 Swiss Francs. The sixth row of Table 5.11 uses the bound of each interval that is closest to zero. For instance, a respondent with a WTP within the interval [2, 3] is assigned a WTP of 2 Swiss Francs. As a result, the means fall to 0.65 or 0.62 Swiss Francs for the weighted and unweighted estimates, respectively. However, this assumptions seems rather strong. To conclude, all sensitivity analyses demonstrate a lower, but still clearly positive WTP. This insight boosts confidence in the robustness of the WTP analysis.

Table 5.12 presents total welfare effects on the benefits side. As only climate externalities are significantly reduced according to my estimates, I only include those. The average CO_2 reduction of 3.5 cents multiplied by 7 gives an average weekly reduction of 0.25 Swiss Francs, and when multiplied by 19 (weeks) an average externality reduction of 4.70 Swiss Francs over the whole treatment period. Average consumer welfare is taken from Table 5.11, and

Table 5.12: Welfare effects of nudging on the benefits side

		Per Report				Whole Period			
		Unweighted		Weighted		Unweighted		Weighted	
			$\sum WB$		$\sum WB$		$\sum WB$		$\sum WB$
Ext. Red.		0.25		0.25		4.70		4.70	
Cons. Welfare	Base case	2.18	2.43	2.42	2.67	41.48	46.13	46.00	50.65
	Uniform WTP at endpoints	1.81	2.06	1.98	2.22	34.45	39.10	37.60	42.26
	WTP={-5,12} at endpoints	1.83	2.08	2.05	2.30	34.77	39.43	39.01	43.66
	WTP={-5,10} at endpoints	1.41	1.65	1.56	1.81	26.70	31.35	29.64	34.30
	WTP={-4,10} at endpoints	1.60	1.84	1.75	2.00	30.31	34.96	33.25	37.91
	WTP bounds closest to 0	0.62	0.87	0.65	0.89	11.78	16.44	12.29	16.95
	Nonresp. have WTP=0	0.96	1.20			18.15	22.80		

Notes: Welfare benefits consist of the sum of monetized external cost savings (depicted in the first row), and consumer welfare changes (depicted in the following rows, including six different welfare robustness estimates). Saved external costs from row 1 are added to the respective consumer welfare estimate to obtain total welfare benefits $\sum WB$ (depicted directly behind the corresponding consumer welfare estimate).

multiplied by 19 to receive the total consumer welfare per average participant over the whole period. Adding the externality reduction of the first row to one of the consumer welfare estimates (including all robustness consumer welfare estimates) leads to the total amount of welfare benefits denoted by $\sum WB$. The weighted base case estimates total weekly benefits at 2.67 Swiss Francs and overall benefits at 50.65 Swiss Francs per average participant. Depending on the assumptions about the WTP endpoints, the overall welfare benefits per participant vary between 16.95 and 50.65 Swiss Francs for weighted estimates.

6 Discussion

The RCT of this thesis used a difference-in-differences design with a control group and showed that nudging significantly reduces climate externalities. The regressions estimate an average reduction of 3.5 cents per day in absolute, or 5.7 percent per day in relative terms. The absolute and relative effect sizes are comparable to results in other nudging experiments. Treatment effects are highly heterogeneous among different groups. People above 55 years, regular car drivers, car owners, biospheric people and people having opened the report email react stronger to the treatment. The total treatment effect on climate costs is additionally significant for people with a household income of between 8'000 and 12'000 Swiss Francs, people in small households, people without a public transport subscription, and for days on weekends.

The ATE on total external costs remains insignificant. One potential reason might be side effects that could not be measured (Tiefenbeck et al., 2013). People may depend too heavily on transportation to show a strong reaction. Instead, they may have tried to reduce their externalities in other areas not included in the study. A second reason might be information overload resulting from various kinds of information in the reports. The statistical power issue involved in this nudging-related RCT due to the limited sample size might be another reason why small effects remain undiscovered.

The analysis of smileys shows significant differences resulting from the kind of feedback participants received in last week's report. People reduce their externalities even further after having received positive feedback and significantly increase their externalities after having received negative feedback. A potential reasoning is that negative feedback reduces the intrinsic motivation of participants (Fong et al., 2019). Although striking, this result is in line with large parts of the psychological literature (Fishbach et al., 2010). The boomerang effect describes unintended consequences of a policy intervention. It may also explain why total externality reductions remain insignificant in this study. Even though the majority of people reduce their externalities, their reduction is outweighed by participants increasing external costs due to negative feedback.
This thesis shows that in contrast to positive feedback leading to expected effects of reducing external costs, negative feedback acts in a counter-productive way by increasing external costs. However, concluding that nudges should only include positive or neutral feedback, and never negative feedback, seems too hasty. According to existing evidence, negative feedback can be motivating as well, at least for certain subgroups (i.e., see Cianci et al. (2010)).

The long-term analysis shows no persisting effect of participants being nudged. As the effect already weakens during the treatment phase, this result is not very surprising. Many other nudging experiments in the energy and environmental fields do not find a long-term effect, either (Ferraro et al., 2011; Brock and Borzino, 2020). Once participants are not repeatedly reminded anymore, the treatment effect disappears or attenuates (Ito et al., 2018).

The WTP analysis estimates that mean WTP lies in the range between 0.62 and 2.42 Swiss Francs per report, even though 41% of respondents revealed a negative WTP. Weighted overall welfare benefits of the intervention lie in the range between 34.30 and 50.65 Swiss Francs per average participant when excluding the implausible estimate where all WTP bounds closest to zero are assigned. Multiplying this amount by approximately 400 participants results in a substantial amount of benefits. If one added private time savings, which were completely neglected here, total benefits would be even higher.

Limitations

The analysis in this thesis contains potential shortcomings, which are briefly discussed in the following. First, some issues related to the study design may question the internal validity of the results. To start, the entire analysis largely depends on the assumptions made about the external costs of the individual transport modes. Moving by bicycle is associated with a questionable positive net external cost of 0.07 Swiss Francs per km. People who wanted to perform well in the study had to deliberately avoid cycling and walk instead, even though cycling is commonly considered a healthy, ecological mean of transportation. In future studies, this should be taken into consideration.

Second, two noteworthy challenges are associated with the data collection via tracking app. Overall, the automatic mode detection works well, but it does not reach a rate of 100%. As only a limited number of all trips were modified and/or confirmed, and not all modes were included in the automatic detection, it may well be that some trips were incorrectly assigned. Moreover, the problem of distinguishing between missing data and zeroes remains an open issue, even though the intensive regressions that were conducted and did not significantly alter the results boost confidence in the internal validity of the results.

Third, the misleading reports during the first six study weeks remain problematic, in particular because all participants began the treatment period on the exact same date. It is impossible to evaluate whether the effect of newly receiving weekly nudges outweighs the wrong content of the reports. Simply excluding the first six weeks may underestimate the average treatment effects, particularly if an accustoming effect prevails. Since the first six weeks should be taken with caution, it is difficult to determine the habituation effect of weekly reports.

The WTP analysis is subject to selection bias, as some of the participants with negative WTP had already opted out of the study before WTP could be elicited in the final survey. Therefore, a robustness check that assigns non-respondents a WTP of zero was conducted. The other estimates potentially overestimate the actual mean WTP.
Moreover, the entire WTP analysis depends on strong assumptions. Setting individual WTP equal to participant welfare requires the assumption that participants can abstract from any behavioral and psychological bias. However, Allcott and Kessler (2019) note that time discounting, habit effects, focusing bias, compromise effects, and the exact timing of asking about WTP, could affect WTP revelation. For instance, participants may underestimate the value of the reports because they filled in the final survey when being annoyed by the last few reports, even though they appreciated the initial reports very much. However, all participants involved in the welfare analysis received several reports, so they know the reports and their content very well and can assess what the reports may bring or harm them in the

future. Therefore, it seems reasonable to assume report receivers are best equipped to know their personal value from weekly reports.

Including all potential costs of nudging into the welfare analysis is essential. Damgaard and Gravert (2018) emphasize that hidden costs of nudges can be problematic when neglected. Bernheim and Taubinsky (2018) criticize nudging as inappropriate when causing a significant number of individuals to experience shame. There is an ongoing debate about the appropriateness of such policies, with some arguing that nudges, which create feelings of shame are an unjustifiable offense to human dignity (Butera et al., 2022). Nudging is also often criticized as being too paternalistic (Tyers, 2019). Goodwin (2012) argues that nudges exploit imperfections in human rationality and are thus manipulative. However, freedom of choice is completely preserved at all stages, and nudging often leads to an increase in welfare. Therefore, if one is aware of the dangers, and designs the nudge accordingly, nudging can be a useful way to effectively and efficiently influence human behavior.

Concerning the external validity of the results, three potential threats come to mind. First, the completely voluntary nature of participation results in self-selection of people who agree to be tracked, and who demonstrate a real interest in this study and related topics. If the willingness to participate is correlated with the response, this selection bias might distort the results. However, at this stage, there is little else to do other than be aware of the potential bias.
Second, this study's population was shown to be more or less comparable to the overall Swiss population. However, due to the focus on German- and French-speaking parts of Switzerland, it is not possible to be completely sure about the population-wide effect.
Finally, the fact that the treatment period starts at the same time for everyone limits the general applicability of the findings. People might have more opportunities and are thus more elastic to reduce their mobility externalities in warmer months, as refraining from using the car is easier then. This would lead to larger treatment effects at other times of the year. Therefore, this study theoretically only allows conclusions about the potential of reducing mobility externalities in the treated months of the year.

Future research
Due to the limited timescale of this thesis, not all analyses could be conducted. In the next step, the following elements are to be investigated further: The WTP analysis could be explored in terms of differences in responses across groups, especially focusing on who is more likely to reveal a negative WTP. Moreover, the welfare analysis is to be completed by including the entire cost side, and private time savings on the benefits side.
To find out whether the treatment effects in the smiley reports are completely exogenous (e.g, whether the estimated treatment effect can be solely attributed to the nudging smiley), one can make use of the six first weeks, in which the reports' contents remained unchanged by accident. This mistake serves as a natural experiment. It allows assigning the varying behavior to either the reports' content or to other circumstances, and thus estimating how large the effects of the smileys really are.

7 Conclusion

The goal of this thesis was to evaluate the treatment and welfare effects of recurring nudges using an RCT in the transport context. The detailed investigation of mobility externalities is of great importance, as forecasts predict an ongoing upward trend in mobility behavior (Statista, 2022).

The thesis estimates that recurring nudges lead to a significant 6%-reduction of daily climate externalities, and a 2%-reduction of daily total external costs, although the latter insignificant. The heterogeneity analysis reveals different treatment effects for various subgroups. Among the ones that respond more strongly to nudging are people above 55 years, people in small households, and people that own or regularly drive a car.
The WTP analysis reveals that participants on average assign a value between 0.62 and 2.42 Swiss Francs to receiving a nudge. The direct benefits to participants thus outweigh the saved climate externalities in absolute terms, which are 0.25 Swiss Francs per nudge on average. Although a precise estimate of total costs created by the nudges is lacking, estimates of the benefits side already provide valuable insights into the implications of nudging.

Acknowledgments

I am very grateful for the valuable support and regular discussions with Prof. Dr. Beat Hintermann who supervised this thesis. I would also like to thank Dr. Beat Schwab, Florian Stich, and Pascal Gächter for their constructive input in the final stage. Moreover, I thank Alexander Götz and Jakob Roth for their valuable input and continuous help throughout the entire process as well as in the final stage.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics*, 95(9-10):1082–1095.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Andor, M. A. and Fels, K. M. (2018). Behavioral economics and energy conservation–a systematic review of non-price interventions and their causal effects. *Ecological economics*, 148:178–210.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Axhausen, K. W., Hintermann, B., Castro, A., Dubernet, T.and Götschi, T., Molley, J., Schoemann, B., Tschervenkov, C., and Tomic, U. (2021). Empirical analysis of mobility behavior in the presence of pigovian transport pricing.

Belschak, F. D. and Den Hartog, D. N. (2009). Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology*, 58(2):274–303.

Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. *Handbook of Behavioral Economics: Applications and Foundations 1*, 1:381–516.

Bonezzi, A., Brendl, C. M., and De Angelis, M. (2011). Stuck in the middle: The psychophysics of goal pursuit. *Psychological science*, 22(5):607–612.

Bothos, E., Mentzas, G., Prost, S., Schrammel, J., and Röderer, K. (2014). Watch your Emissions: Persuasive Strategies and Choice Architecture for Sustainable Decisions in Urban Mobility. *PsychNology Journal*, 12(3).

Brock, M. and Borzino, N. (2020). Using incentives and social information to promote energy conservation behavior: Field experiment. *Sustainability and Environmental Decision Making*, pages 1–22.

Butera, L., Metcalfe, R., Morrison, W., and Taubinsky, D. (2022). Measuring the welfare effects of shame and pride. *American Economic Review*, 112(1):122–68.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Carreras, I., Gabrielli, S., Miorandi, D., Tamilin, A., Cartolano, F., Jakob, M., and Marzorati, S. (2012). SUPERHUB: a user-centric perspective on sustainable urban mobility. In *Proceedings of the 6th ACM workshop on Next generation mobile computing for dynamic personalised travel planning*, pages 9–10.

Cellina, F., Bucher, D., Mangili, F., Veiga Simão, J., Rudel, R., and Raubal, M. (2019). A large scale, app-based behaviour change experiment persuading sustainable mobility patterns: Methods, results and lessons learnt. *Sustainability*, 11(9):2674.

Cianci, A. M., Klein, H. J., and Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology*, 95(4):618.

Dai, H., Saccardo, S., Han, M. A., Roh, L., Raja, N., Vangala, S., Modi, H., Pandya, S., Sloyan, M., and Croymans, D. M. (2021). Behavioural nudges increase covid-19 vaccinations. *Nature*, 597(7876):404–409.

Damgaard, M. T. and Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157:15–26.

Delft, C. (2019). Handbook on the external costs of transport: Version 2019. *Europese Commissie, publicatienummer*, 18.

Delmas, M. A., Fischlein, M., and Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, 61:729–739.

Federal Roads Office - ASTRA (2017). Handbook NISTRA 2017. `https://www.astra.admin.ch/astra/de/home/fachleute/dokumente-nationalstrassen/fachdokumente/nistra.html` Last accessed: 07/06/2022.

Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–22.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Fishbach, A., Eyal, T., and Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. *Social and Personality Psychology Compass*, 4(8):517–530.

Fong, C. J., Patall, E. A., Vasquez, A. C., and Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review*, 31(1):121–162.

Goodwin, T. (2012). Why we should reject 'nudge'. *Politics*, 32(2):85–92.

Gravert, C. and Collentine, L. O. (2021). When nudges aren't enough: Norms, incentives and habit formation in public transport usage. *Journal of Economic Behavior & Organization*, 190:1–14.

Hintermann, B., Schoeman, B., Molloy, J., Götschli, T., Castro, A., Tchervenkov, C., Tomic, U., and Axhausen, K. W. (2021). Pigovian transport pricing in practice.

Hülsmann, F., Gerike, R., Kickhöfer, B., Nagel, K., and Luz, R. (2011). *Towards a multi-agent based modeling approach for air pollutants in urban regions.*

Hummel, D. and Maedche, A. (2019). How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80:47–58.

Ito, K., Ida, T., and Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1):240–67.

Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., and Walker, J. L. (2015). Quantified traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems*, 19(2):109–124.

Karich, P. and Schröder, S. (2014). Graphhopper. *http://www. graphhopper. com, last accessed on 27 May 2022*, 4(2):15.

Kickhöfer, B., Hülsmann, F., Gerike, R., and Nagel, K. (2013). Rising car user costs: comparing aggregated and geo-spatial impacts on travel demand and air pollutant emissions. In *Smart Transport NEtworks*. Edward Elgar Publishing.

Kotz, S. and Van Dorp, J. R. (2004). *Beyond beta: other continuous families of distributions with bounded support and applications.* World Scientific.

Kristal, A. S. and Whillans, A. V. (2020). What we can learn from five naturalistic field experiments that failed to shift commuter behaviour. *Nature Human Behaviour*, 4(2):169–176.

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends® in Econometrics*, 4(3):165–224.

Löschel, A., Rodemeier, M., and Werthschulte, M. (2020). When nudges fail to scale: field experimental evidence from goal setting on mobile phones. *ZEW-Centre for European Economic Research Discussion Paper*, (20-039).

Macours, K. and Molina Millán, T. (2017). Attrition in randomized control trials: Using tracking information to correct bias.

Maerivoet, S., Daems, F., Maertens, F., Renckens, K., Van Houtte, P., and Buelens, L. (2012). A field trial on smart mobility. *Procedia-Social and Behavioral Sciences*, 54:926–935.

Marangoni, G. and Tavoni, M. (2021). Real-time feedback on electricity consumption: evidence from a field experiment in italy. *Energy Efficiency*, 14(1):1–17.

McKerracher, C. and Torriti, J. (2013). Energy consumption feedback in perspective: integrating australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2):387–405.

Molloy, J., Schatzmann, T., Schoeman, B., Tchervenkov, C., Hintermann, B., and Axhausen, K. W. (2021). Observed impacts of the covid-19 first wave on travel behaviour in switzerland based on a large gps panel. *Transport Policy*, 104:43–51.

Möser, G. and Bamberg, S. (2008). The effectiveness of soft transport policy measures: A critical assessment and meta-analysis of empirical evidence. *Journal of Environmental Psychology*, 28(1):10–26.

Myers, E. and Souza, M. (2020). Social comparison nudges without monetary incentives: Evidence from home energy reports. *Journal of Environmental Economics and Management*, 101:102315.

Nikolic, M. and Bierlaire, M. (2017). Review of transportation mode detection approaches based on smartphone data. In *17th Swiss Transport Research Conference*, number CONF.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and social psychology bulletin*, 34(7):913–923.

Pluntke, C. and Prabhakar, B. (2013). Insinc: a platform for managing peak demand in public transit. *JOURNEYS, Land Transport Authority Academy of Singapore*, 2013:31–39.

Rela, N. (2022). On nudges that fail. *Behavioural Public Policy*, pages 1–15.

Rosenfield, A., Attanucci, J. P., and Zhao, J. (2020). A randomized controlled trial in travel demand management. *Transportation*, 47(4):1907–1932.

Sasaki, S., Saito, T., and Ohtake, F. (2022). Nudges for covid-19 voluntary vaccination: How to explain peer information? *Social Science & Medicine*, 292:114561.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier.

Statista (2022). Statistiken zur Mobilität in der Schweiz. `https://de.statista.com/themen/2031/mobilitaet-in-der-schweiz/#dossierKeyfigures` Last accessed: 07/06/2022.

Steg, L. and De Groot, J. (2010). Explaining prosocial intentions: Testing causal relationships in the norm activation model. *British journal of social psychology*, 49(4):725–743.

Swiss Federal Statistical Office (2018). Statistischer Bericht 2018. Technical report, Swiss Federal Statistical Office, Mobility Section.

Thaler, R. and Sunstein, C. (2008). *Nudge: improving decisions about health, wealth and happiness Penguin*. Penguin Books, New York.

Thunström, L. (2019). Welfare effects of nudges: The emotional tax of calorie menu labeling. *Judgment and Decision making*, 14(1):11.

Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013). For better or for worse? empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160–171.

Tyers, R. (2018). Nudging the jetset to offset: voluntary carbon offsetting and the limits to nudging. *Journal of Sustainable Tourism*, 26(10):1668–1686.

Tyers, R. (2019). Macro libertarianism and micro paternalism: Governance in an age of nudging. *Handbook of Behavioural Change and Public Policy*, pages 332–343.

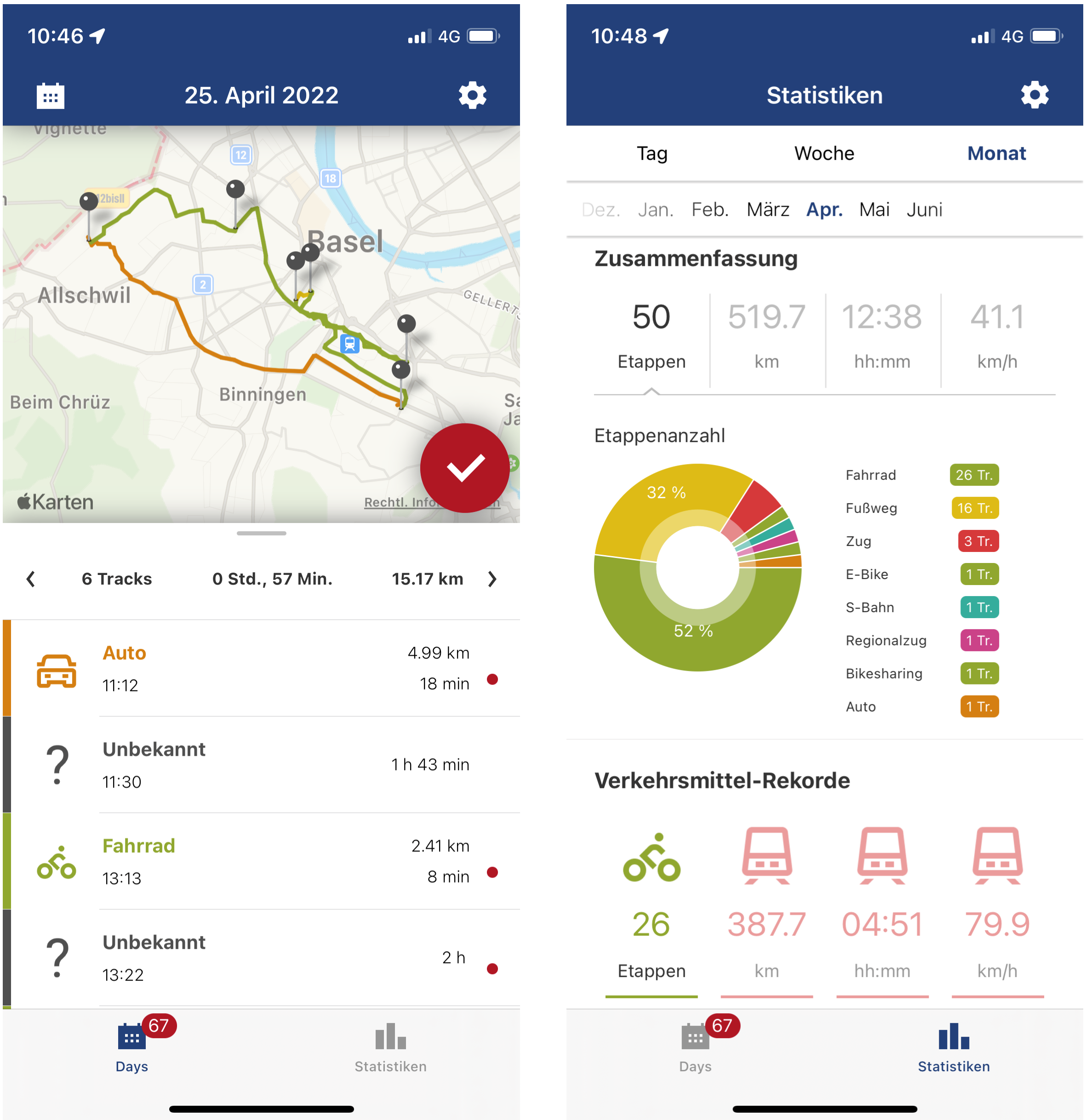
Verhoef, E. T. (2000). The implementation of marginal external cost pricing in road transport. *Papers in Regional Science*, 79(3):307–332.

Wilson, N. L., Just, D. R., Swigert, J., and Wansink, B. (2017). Food pantry selection solutions: a randomized controlled trial in client-choice food pantries to nudge clients to targeted foods. *Journal of Public Health*, 39(2):366–372.

Wu, L., Yang, B., and Jing, P. (2016). Travel mode detection based on gps raw data collected by smartphones: a systematic review of the existing methodologies. *Information*, 7(4):67.

A Appendix

Figure A.1: Example of the GPS-tracking app “Catch-my-Day”



Note: Example tracks on a given day in April 2022 are depicted. Participants could correct the automatic mode detection by clicking on the respective track. By ticking the red check mark, participants could confirm all recorded tracks of a certain day.

Table A.1: Daily common trends regression, per mode of transport

	(1) Total Ext.	(2) Total Ext.	(3) Total Car Ext.	(4) Total PT Ext.	(5) Total Bike Ext.	(6) Total Walk Ext.
Treatment	0.284+	-8.161	-28.058	14.042+	1.379	-0.599
	(0.171)	(45.522)	(46.285)	(8.276)	(3.001)	(2.704)
Date	0.000	-0.000	-0.001	0.001***	-0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
Treatment x Date		0.000	0.001	-0.001+	-0.000	0.000
		(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
adj. R^2	0.00062	0.00061	0.00088	0.00089	0.00085	0.00021
Clusters	777	777	777	777	777	777
N	69219	69219	69219	69219	69219	69219

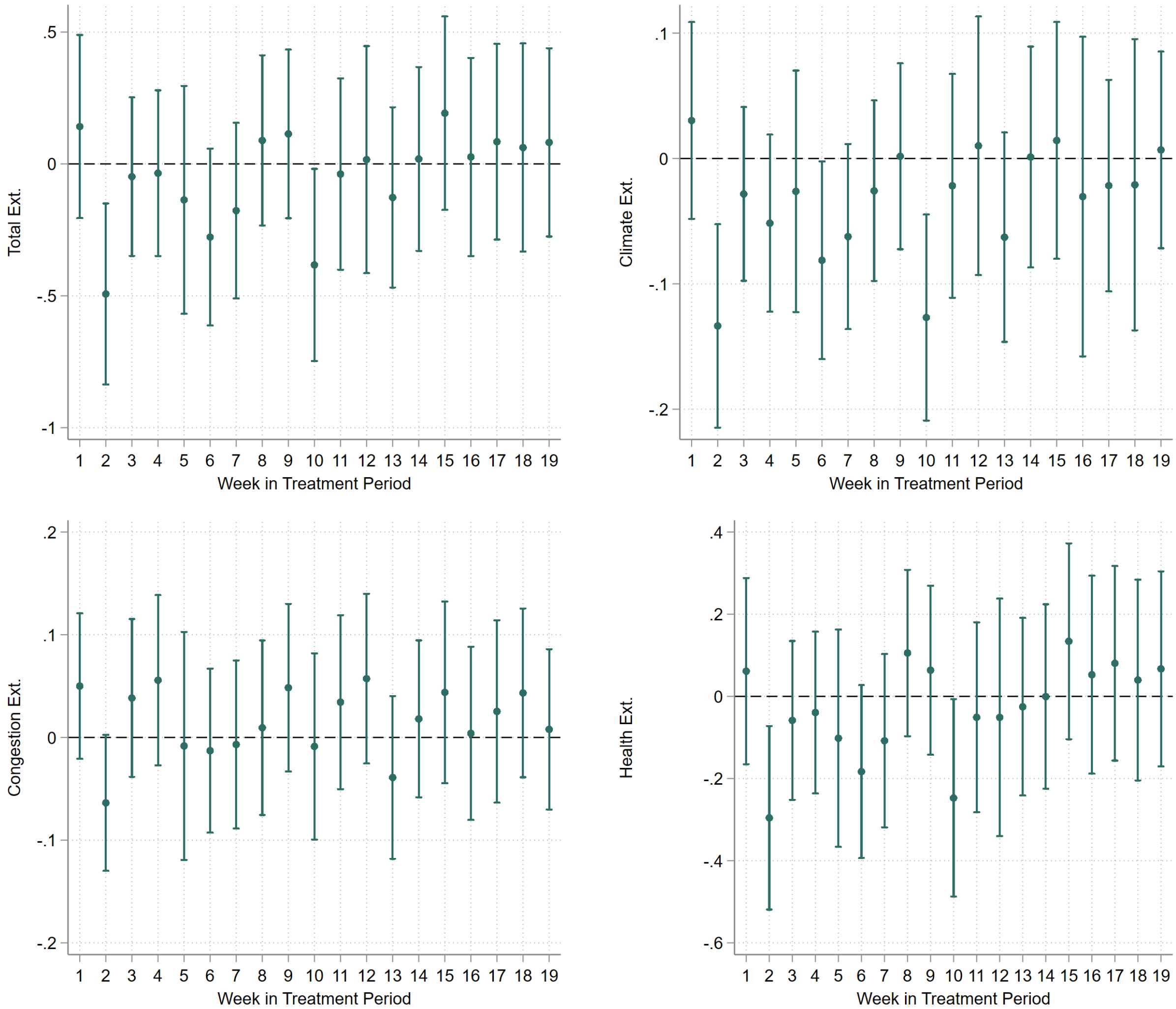
Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). Regressions are run on pre-treatment averages, only including observations up to the beginning of the treatment period. The dependent variable is aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. No fixed effects are included.

Table A.2: Probit regressions to estimate the propensity to finish the study

	(1) P (Finish treatment period=1)	(2) P (Finish post-treatment period=1)
Treatment Group	0.047	-0.042
	(0.081)	(0.081)
Male	0.102	0.037
	(0.091)	(0.091)
Age	0.013^{***}	0.020^{***}
	(0.004)	(0.004)
Higher education	0.168^{*}	0.136
	(0.085)	(0.085)
Household income	-0.061	-0.050
	(0.033)	(0.033)
Household size	-0.025	-0.037
	(0.035)	(0.035)
French	-0.062	-0.015
	(0.107)	(0.107)
Fulltime	0.035	0.125
	(0.097)	(0.097)
Retired	-0.121	-0.238
	(0.176)	(0.174)
Reg. car	-0.116	-0.128
	(0.094)	(0.095)
Reg. bike	0.012	-0.015
	(0.129)	(0.128)
Reg. PT	-0.060	-0.073
	(0.107)	(0.107)
Constant	-0.271	-0.788^{***}
	(0.218)	(0.220)
N	1009	1009
AIC	1359.4	1372.1
BIC	1423.3	1436.0

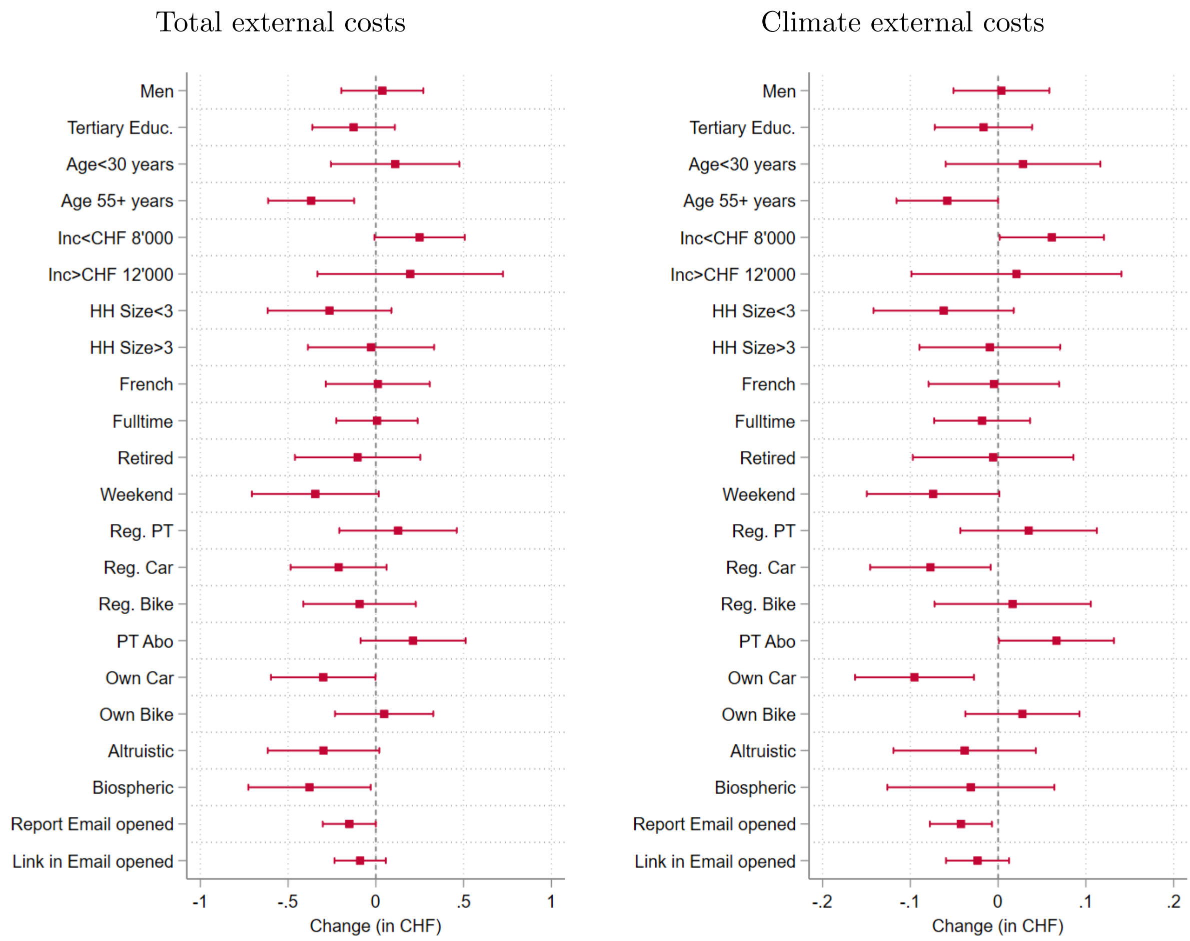
Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is a dummy equal to 1 if a participant was still tracking within the last two study weeks (of the treatment or post-treatment period in columns 1 and 2, respectively), and 0 otherwise. The regressions include all 1009 participants that were part of the study at the beginning of the observation period. Robust standard errors in parentheses.

Figure A.2: Event study approach



Note: The Figure estimates separate (non-linear) total treatment effects for each study week of the treatment period. The average over the pre-treatment period serves as reference category (instead of the last week before treatment begins). The DiD estimates include date and participant fixed effects. Standard errors are clustered at the participant level. 95%-confidence intervals are shown.

Figure A.3: Treatment effect heterogeneity: Interaction terms



Note: Only interaction terms of univariate regressions including one type of interaction (i.e., household income) at a time are depicted. 95%-confidence intervals are shown.

Table A.3: Effect heterogeneity: Gender

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.079	-0.037	-0.006	-0.036
	(0.104)	(0.023)	(0.030)	(0.064)
Male	0.037	0.004	0.036	-0.003
	(0.119)	(0.028)	(0.032)	(0.075)
Treated + Treated x Male	-0.042	-0.033	0.030	-0.039
	(0.102)	(0.024)	(0.025)	(0.065)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.4: Effect heterogeneity: Education

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.004	-0.027	0.043+	-0.012
	(0.107)	(0.025)	(0.025)	(0.069)
Tertiary Educ.	-0.127	-0.017	-0.058+	-0.053
	(0.120)	(0.028)	(0.031)	(0.075)
Treated + Treated x Tertiary Educ.	-0.123	-0.043	-0.014	-0.065
	(0.099)	(0.023)	(0.029)	(0.061)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.5: Effect heterogeneity: Age

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.056	-0.019	0.054+	0.021
	(0.098)	(0.022)	(0.030)	(0.061)
Old	-0.369**	-0.058*	-0.088**	-0.223**
	(0.125)	(0.029)	(0.033)	(0.080)
Young	0.110	0.028	-0.035	0.117
	(0.186)	(0.045)	(0.045)	(0.113)
Treated + Treated x Old	-0.313**	-0.077**	-0.034	-0.201**
	(0.113)	(0.027)	(0.026)	(0.073)
Treated + Treated x Young	0.166	0.009	0.019	0.138
	(0.179)	(0.043)	(0.040)	(0.109)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.6: Effect heterogeneity: Household income

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.158+	-0.057*	-0.002	-0.098+
	(0.092)	(0.022)	(0.025)	(0.058)
Treated x High Inc.	0.196	0.021	0.062	0.113
	(0.269)	(0.061)	(0.081)	(0.154)
Treated x Low Inc.	0.249+	0.061*	0.035	0.153+
	(0.131)	(0.030)	(0.031)	(0.084)
Treated + Treated x High Inc.	0.038	-0.036	0.060	0.015
	(0.266)	(0.060)	(0.081)	(0.152)
Treated + Treated x Low Inc.	0.091	0.004	0.033	0.055
	(0.125)	(0.028)	(0.028)	(0.080)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.7: Effect heterogeneity: Household size

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.100	0.003	0.051	0.047
	(0.169)	(0.038)	(0.042)	(0.104)
Treated x Large HH size	-0.027	-0.009	-0.030	0.012
	(0.183)	(0.041)	(0.046)	(0.113)
Treated x Small HH size	-0.264	-0.062	-0.048	-0.153
	(0.180)	(0.041)	(0.044)	(0.111)
Treated + Treated x Large HH size	0.073	-0.007	0.021	0.059
	(0.109)	(0.025)	(0.030)	(0.068)
Treated + Treated x Small HH size	-0.163	-0.059*	0.003	-0.107
	(0.103)	(0.024))	(0.027)	(0.065)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.8: Effect heterogeneity: Language

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.060	-0.034+	0.014	-0.039
	(0.089)	(0.021)	(0.024)	(0.056)
Treated x French	0.011	-0.005	0.006	0.010
	(0.151)	(0.038)	(0.033)	(0.096)
Treated + Treated x French	-0.049	–0.039	0.020	-0.029
	(0.148)	(0.037)	(0.032)	(0.094)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.9: Effect heterogeneity: Fulltime

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.061	-0.026	-0.024	-0.012
	(0.091)	(0.021)	(0.022)	(0.058)
Fulltime=1	0.006	-0.018	0.074*	-0.050
	(0.118)	(0.028)	(0.030)	(0.074)
Treated + Treated x Fulltime	-0.055	-0.044+	0.050+	-0.061
	(0.112)	(0.026)	(0.030)	(0.070)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.10: Effect heterogeneity: Retired

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.048	-0.035+	0.022	-0.035
	(0.087)	(0.020)	(0.023)	(0.055)
Treated x Retired	-0.104	-0.006	-0.074	-0.024
	(0.182)	(0.047)	(0.048)	(0.104)
Treated + Treated x Retired	-0.152	-0.040	-0.052	-0.059
	(0.180)	(0.046)	(0.048)	(0.103)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.11: Effect heterogeneity: Weekend

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.041	-0.014	0.024	0.030
	(0.099)	(0.021)	(0.028)	(0.061)
Treated x Weekend	-0.345+	-0.074+	-0.033	-0.238+
	(0.184)	(0.038)	(0.039)	(0.122)
Treated + Treated x Weekend	-0.304+	-0.088*	-0.008	-0.208*
	(0.156)	(0.036)	(0.028)	(0.105)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.12: Effect heterogeneity: Regular public transport user

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.080	-0.041*	0.013	-0.051
	(0.088)	(0.020)	(0.023)	(0.055)
Treated x Reg. PT	0.127	0.035	0.012	0.080
	(0.171)	(0.040)	(0.038)	(0.107)
Treated + Treated x Reg. PT	0.047	-0.006	0.025	0.029
	(0.168)	(0.039)	(0.037)	(0.106)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.13: Effect heterogeneity: Regular car user

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.002	-0.015	0.014	-0.001
	(0.091)	(0.020)	(0.025)	(0.057)
Treated x Reg. car	-0.212	-0.077*	0.004	-0.139
	(0.139)	(0.035)	(0.031)	(0.089)
Treated + Treated x Reg. car	-0.214	-0.092**	0.018	-0.140
	(0.134)	(0.034)	(0.029)	(0.086)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.14: Effect heterogeneity: Regular bike user

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.047	-0.037+	0.017	-0.027
	(0.088)	(0.020)	(0.023)	(0.055)
Treated x Reg. bike	-0.093	0.017	-0.016	-0.093
	(0.163)	(0.045)	(0.032)	(0.116)
Treated + Treated x Reg. bike	-0.140	-0.020	0.001	-0.120
	(0.161)	(0.045)	(0.031)	(0.115)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.15: Effect heterogeneity: PT subscription

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.101	-0.049*	0.019	-0.071
	(0.089)	(0.021)	(0.024)	(0.056)
Treated x Abo	0.212	0.066*	-0.019	0.164+
	(0.152)	(0.033)	(0.036)	(0.095)
Treated + Treated x Abo	-0.111	0.018	-0.000	0.094
	(0.149)	(0.032)	(0.035)	(0.093)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.16: Effect heterogeneity: Own car

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.178	0.040	0.000	0.138
	(0.148)	(0.034)	(0.030)	(0.094)
Treated x Own car	-0.300*	-0.095**	0.019	-0.223*
	(0.152)	(0.034)	(0.032)	(0.096)
Treated + Treated x Own car	-0.121	-0.055**	0.019	-0.085
	(0.089)	(0.021)	(0.024)	(0.056)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.17: Effect heterogeneity: Own bike

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.091	-0.055+	0.029	-0.065
	(0.139)	(0.032)	(0.036)	(0.084)
Treated x Own bike	0.047	0.028	-0.019	0.039
	(0.143)	(0.033)	(0.037)	(0.087)
Treated + Treated x Own bike	-0.044	-0.027	0.009	-0.026
	(0.089)	(0.021)	(0.023)	(0.057)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.18: Effect heterogeneity: Altruistic

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.077	-0.028	0.078*	0.027
	(0.151)	(0.039)	(0.038)	(0.095)
Treated x Altruistic	-0.298+	-0.038	-0.103*	-0.157
	(0.162)	(0.041)	(0.042)	(0.102)
Treated + Treated x Altruistic	-0.221+	-0.066*	-0.025	-0.130+
	(0.116)	(0.027)	(0.032)	(0.073)
adj. R^2	0.179	0.166	0.143	0.186
Clusters	483	483	483	483
N	94232	94232	94232	94232

Table A.19: Effect heterogeneity: Biospheric

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.152	-0.032	0.092*	0.092
	(0.171)	(0.047)	(0.042)	(0.106)
Treated x Biospheric	-0.378*	-0.031	-0.113*	-0.234*
	(0.177)	(0.048)	(0.045)	(0.111)
Treated + Treated x Biospheric	-0.226*	-0.063*	-0.020	-0.142*
	(0.111)	(0.026)	(0.030)	(0.070)
adj. R^2	0.182	0.168	0.144	0.189
Clusters	483	483	483	483
N	94203	94203	94203	94203

Table A.20: Effect heterogeneity: Report opened

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.031	-0.010	0.020	0.021
	(0.098)	(0.022)	(0.025)	(0.061)
Treated x Report opened	-0.151*	-0.042*	-0.009	-0.100*
	(0.077)	(0.018)	(0.019)	(0.048)
Treated + Treated x Report opened	-0.120	-0.053*	0.011	-0.078
	(0.088)	(0.021)	(0.023)	(0.056)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.21: Effect heterogeneity: Link in report opened

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	-0.029	-0.028	0.015	-0.017
	(0.088)	(0.020)	(0.023)	(0.055)
Treated x Link in report opened	-0.090	-0.023	-0.001	-0.065
	(0.074)	(0.018)	(0.018)	(0.048)
Treated + Treated x Link in report opened	-0.118	-0.051*	0.014	-0.082
	(0.097)	(0.023)	(0.024)	(0.062)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Table A.22: Effect heterogeneity: Weekdays

	(1) Total Ext.	(2) Climate Ext.	(3) Cong. Ext.	(4) Health Ext.
Treated	0.103	-0.004	0.039	0.069
	(0.129)	(0.027)	(0.036)	(0.081)
Treated x Tuesday	-0.163	-0.029	-0.049	-0.084
	(0.136)	(0.030)	(0.036)	(0.087)
Treated x Wednesday	-0.118	-0.020	-0.016	-0.082
	(0.142)	(0.032)	(0.036)	(0.091)
Treated x Thursday	-0.114	-0.015	-0.050	-0.049
	(0.150)	(0.035)	(0.039)	(0.095)
Treated x Friday	0.054	0.012	0.037	0.006
	(0.157)	(0.035)	(0.036)	(0.100)
Treated x Saturday	-0.306	-0.051	-0.031	-0.224
	(0.226)	(0.047)	(0.046)	(0.151)
Treated x Sunday	-0.500*	-0.116*	-0.058	-0.326*
	(0.228)	(0.048)	(0.048)	(0.150)
adj. R^2	0.176	0.168	0.150	0.181
Clusters	777	777	777	777
N	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. The reference category contains treated observations on Mondays. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Table A.23: Average treatment effects on climate externalities with static smiley effects

	(1) Clim. Ext.	(2) Clim. Ext.	(3) Clim. Ext.	(4) Clim. Ext.	(5) Clim. Ext.	(6) Clim. Ext.	(7) Clim. Ext.
Treated	-0.035+	0.013	0.018	-0.071***	-0.069***	-0.025	-0.033+
	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)
Treated x ☺(☺)		-0.229***					
		(0.021)					
Treated x ☺ (Base)			-0.232***			-0.171***	
			(0.018)			(0.018)	
Treated x ☹(☹)				0.614***			
				(0.060)			
Treated x ☹ (Base)					0.447***	0.386***	
					(0.034)	(0.034)	
Treated x ☺☺							-0.236***
							(0.021)
Treated x ☺							-0.072*
							(0.029)
Treated x ☹							0.385
							(0.257)
Treated x ☹☹							0.602***
							(0.041)
adj. R^2	0.168	0.169	0.169	0.171	0.170	0.171	0.172
Clusters	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Table A.24: Average treatment effects on climate externalities with dynamic smiley effects

	(1) Clim. Ext.	(2) Clim. Ext.	(3) Clim. Ext.	(4) Clim. Ext.	(5) Clim. Ext.	(6) Clim. Ext.	(7) Clim. Ext.
Treated	-0.035+	-0.031	-0.031	-0.030	-0.042*	-0.054**	-0.055**
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Treated x (☹☹ to ☺☺)		-0.316***					
		(0.038)					
Treated x (☹(☹) to ☺(☺))			-0.141***				
			(0.040)				
Treated x (☹ to ☺ (Base))				-0.105***			
				(0.030)			
Treated x (☺☺ to ☹☹)					0.577***		
					(0.076)		
Treated x (☺(☺) to ☹(☹))						0.597***	
						(0.065)	
Treated x (☺ to ☹ (Base))							0.412***
							(0.037)
adj. R^2	0.168	0.168	0.168	0.168	0.168	0.170	0.169
Clusters	777	777	777	777	777	777	777
N	145569	145569	145569	145569	145569	145569	145569

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable is the external cost of transport aggregated to the person-day level. Standard errors (in parentheses) are clustered at the participant level. All regressions include date and person fixed effects.

Figure A.4: Tracking statistics including post-treatment period

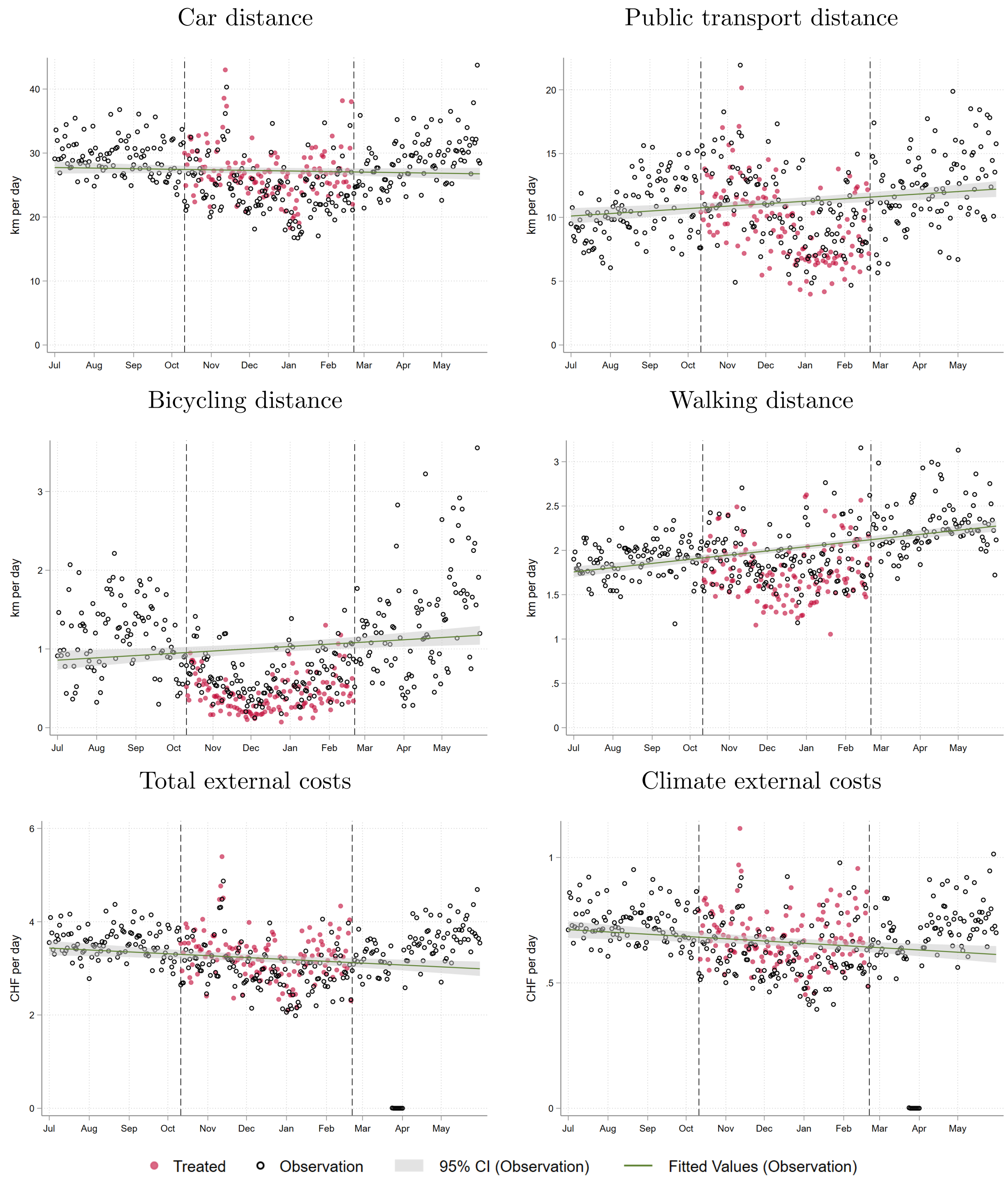


Table A.25: WTP Analysis: Probit regression for inverse probability weighting

	P (Final survey=1)
Male	0.411*
	(0.184)
Age	-0.011
	(0.007)
Higher education	0.135
	(0.153)
HH Income preferred not to say	*Ref.*
HH Income 4'000 CHF or less	-0.458
	(0.479)
HH Income 4'001 - 8'000 CHF	-0.227
	(0.313)
HH Income 8'001 - 12'000 CHF	0.107
	(0.285)
HH Income 12'001 - 16'000 CHF	-0.251
	(0.327)
HH Income 16'001 CHF or more	-0.085
	(0.338)
Household size	-0.205**
	(0.071)
French	-0.231
	(0.202)
English	0.001
	(0.447)
Fulltime	-0.158
	(0.190)
Retired	-0.989**
	(0.325)
Living in partnership	1.769***
	(0.173)
Reg. car	-0.288
	(0.241)
Reg. bike	-0.148
	(0.268)
Reg. PT	-0.073
	(0.240)
Mobis participant	-0.624*
	(0.265)
Share of reports opened	0.099
	(0.303)
Share of links in reports opened	0.686
	(0.364)
Constant	0.148
	(0.514)
N	409
AIC	491.7
BIC	407.4

Notes: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (based on two-sided testing). The dependent variable equals 1 if the participant answered the WTP questions in the final survey, and 0 otherwise. Robust standard errors in parentheses.