FIT Working Paper 16

Lassi Ahlvik \& Anna Sahari

## Promoting active transport through health information: evidence from a randomized controlled trial

# Promoting active transport through health information: evidence from a randomized controlled trial 

Lassi Ahlvik, Anna Sahari*

December 18, 2023


#### Abstract

The negative health impacts of passive transport can be substantial both from the individual point of view and with respect to public spending on health. We run a fieldexperiment to study the effectiveness of information in encouraging active transport choices. The treatment is carried out using a smartphone application that automatically tracks users' daily transport choices. We find that the treatment group receives and internalizes the information, but we do not find a statistically significant effect on their travel behavior.


Keywords: Transport Externalities, Active travel, Health Behaviour, Nudging
JEL codes: I12, R41, D90

[^0]
## 1 Introduction

Insufficient physical activity, sometimes referred to as an 'inactivity pandemic', is the fourth largest risk factor for mortality and, as such, incurs significant economic costs for individuals (Kohl et al., 2012). ${ }^{1}$ How should public policy react to these massive costs? There are two alternative approaches. If costs are fully internalized, lack of activity can be seen as a rational choice and private health costs do not justify governmental intervention. Regulation is only aimed to target the fiscal externalities, stemming from the fact that individuals do not bear the full cost of health care services. If costs are not internalized, however, individual costs can directly justify a public policy intervention. The required intervention would be sizable: it has been shown that the adverse health effects from driving surpass other externalities, and the corrective fuel taxes would be up to $50 \%$ higher than those without individual health considerations (van den Bijgaart et al., 2020). ${ }^{2}$ And if behavioral biases exist, 'nudges' could alter individuals' actions at low cost.

In this study we design and run an informational randomized controlled trial (RCT) to shed light on this issue. Individuals may not internalize the health effects because they hold biased information about the health effects of inactivity, the recommendations for activity levels or the activity of their peers. If providing individuals with this information changes their transport behavior there is a behavioral bias that public policy should correct. It is usually difficult to observe actual transport behavior at the individual level. We overcome this problem by using a mobile phone application with an algorithm that tracks individuallevel mobility and automatically recognizes the transport mode. The application allows us to target personalized health-related information about mobility choices to the treatment group.

We make three main findings. First, the information reached the treatment group, as the daily interactions with the application grew by $51 \%$. The increase was the largest

[^1]on Mondays when we sent informational push-notifications, but engagement stayed high also on other days of the week. Second, we find evidence that the users internalized the information, at least to some extent, because we find a positive treatment effect on healthrelated knowledge in the end-of-treatment survey. Third, we find no statistically significant effect on transport choices, whether measured by walking, cycling or public transit use.

Our results contribute to the literature on to what extent individuals internalize the health effects of their transport choices. The literature generally finds that individuals have wrong beliefs about the the probability of disease associated with physical inactivity (Haase et al., 2004; Coups et al., 2008; Bennett et al., 2009). Our study finds support for these findings and shows that wrong beliefs can be corrected. Our contribution is to go beyond survey evidence because we observe actual behavior. We do not find evidence of information avoidance as described in Golman et al. (2017), but we nevertheless observe no change in individual-level travel behavior. Our findings can be interpreted as support for the "rational choice" model of transport behavior.

This study also contributes to the literature on the effects of information treatments and nudges. Within the context of transport, data in these studies often measure public transport or car use, as these are easier to observe than walking or cycling. Recent examples of randomized experiments include Gravert and Olsson Collentine (2021) on the effect of social comparisons on public transport use, Lieberoth et al. (2018) on the effect of health information on public transport use, Kristal and Whillans (2020) on the effect of different information interventions on car use and Gessner et al. (2023) and Goetz et al. (2023) on the effect of social comparisons and moral appeals on mode choice. A general finding in this literature is that nudges are not effective in changing transportation habits. Only Gessner et al. (2023) find that information provision impacts transport choices: their treatment combining both moral appeals and social comparisons results in a reduction of $38 \%$ in spending on car-related mobility services in the context of a transport budget provided by the employer.

Our study adds to this literature in three ways. First, the mobile phone interface allows us to send personalized health information to our treatment group. Since we observe daily transport choices, we can condition the health information and informational messages on the individual's observed behaviour during the previous week. Second, with the mobile application, we can verify that information reached the treatment group. This allows us to rule out a hypothesis that the null result is simply caused by no-one receiving the information treatment. Third, we track individual's actual travel behavior over a relatively long period of
three months. We observe these data for all transport modes, including walking and cycling for which the data are usually self-reported and collected by surveys. Such observational data are still rare, although similar mobile tracking has been used in recent work by Cellina et al. (2019), Goetz et al. (2023), Schwab et al. (2023) and Hintermann et al. (2022). ${ }^{3}$ Out of these studies, the experiment by Goetz et al. (2023) is closest to ours in terms of treatment design, but their goal is to reduce the climate impact of transport. Similar to our findings, they observe no effect on transport behaviour.

## 2 Data and experimental design

### 2.1 Data sources

Mobile phone application for individual-level mobility data. The main data source is a carbon footprint calculator embedded in a mobile phone application. The main application, "Tampere.finland", is provided by the city of Tampere and it contains for example news, events, public transit routes and timetables. One feature of the application is the carbon footprint calculator. When a user turns on the calculator, it automatically tracks mobility and detects the transport mode of each trip. ${ }^{4}$ The user can correct potential mistakes manually during three days, after which the trip is locked. Based on this information, the calculator produces a carbon footprint for each trip and tracks the total carbon footprint on a monthly level. This application was developed by the city of Tampere, and the researchers and developers jointly designed the treatment analyzed in this paper.

Our analysis focuses on residents of Pirkanmaa region, which includes the city of Tampere and 22 other municipalities. ${ }^{5}$ Although the application is accessible to users throughout Finland, the algorithm can accurately recognize municipal public transportation only in this region. We have access to daily trip data since the carbon footprint calculator was introduced on June 1st, 2021. As a result, we can observe the application's users for at most one year leading up to the three-month intervention, which started on September 1st, 2022. The

[^2]calculator had 1,442 active users in August 2022 and it has had 5,494 unique users (device identification numbers) between June 1st 2021 and August 30th 2022.

The data consist of daily observations of the total length (in meters) and time (in minutes) travelled using each transport mode (walking, cycling, bus, tram, train and car) as well as the total number of trips per day, identified by a unique device identification number. The device id. allows us to track the same individuals over the entire study period unless they change their mobile phone or re-install the application. We have access to the postal code of the starting point of the first trip of the day. We interpret the most commonly observed starting point as the user's residential location. To measure how the information reaches the end-users, we employ data on daily application use captured by the application programming interfaces (API) calls per user per day, which are collected from August 10th 2022 onward.

We complement these base data with open access postal code -level data from different sources: Socio-demographic information available at Statistics Finland Paavo-database and the availability of public transport accessed through the city of Tampere.

Representativeness. All users of the carbon footprint application were included in the experiment (opt-out setting), but using the application was voluntary. We explore the representativeness of our sample in two ways. First, to assess geographic representativeness, we connect the application users to each postal code in the Pirkanmaa region based on the postal code where the first recorded trip of each day begins. Figure A. 2 in Appendix A reveals that our sample contains observations also from sparsely populated areas, and the number of observations per postal code rises proportionally with the total postal code population level.

Second, to study whether application users behave systematically differently from the general population, Table A. 1 compares our mobile phone -based data to the National Travel Survey's subsample representing the Pirkanmaa region, carried out in 2016 and 2021. In terms of length, our data are consistent with the survey data, including somewhat more walking and less public transport use. In terms of time, the application gives longer duration for walking, as the application counts the entire trip duration to its length (e.g. a visit to the store). For other modes of transport, the timing matches the survey responses well.

Information survey. To find out whether the individuals understand and internalize the information provided in the treatment, we ran an end-of-treatment online survey between December 7 and December 30 2022. We invited application users by a push notification
sent to their phones, linking to an external web page where the survey was located. ${ }^{6}$ We incentivized people to respond by giving out nine complimentary movie tickets allocated randomly to respondents who provided their e-mail address. The total number of responses was 141 and the response rate was thus $9.8 \%$.

The aim was to explore whether treatment and control groups respond differently to survey questions, which measure their knowledge about active transport and health. We asked two questions to identify their level of knowledge. First, " According to your assessment, what share of Finns aged over 30 commute to work by cycling or walking?" and, second, " According to your assessment, how much does regular, sufficient activity reduce the overall risk of mortality? ( $0-10 \%, 10-20 \%, 20-30 \%, 30-40 \%, 40-50 \%)(0-20 \%, 20-40 \%, 40-60 \%, 60-$ $80 \%, 80-100 \%)$. In addition, we asked to report the age, gender, and presence of underaged children in the households of the respondents and used them as covariates in the analysis.

### 2.2 Experimental design

During the intervention, the application users were randomized into treatment and control groups. ${ }^{7}$ The control group continued to use the standard version of the carbon footprint calculator. The treatment group had additional elements in the calculator, with visuals that communicated the personalized amounts of walking and cycling on a daily/weekly/monthly level, whether the individual had reached the WHO's goal for weekly activity, social comparison among the application users and personalized health impacts based on the recorded activity.

Figure 1 shows all changes to the application for the treatment group. The first new view showed the total weekly minutes of walking and cycling, and how they contribute to the achievement of the WHO recommendation of 150 active minutes per week (Figure 1a). The view also shows information on historical activity on weekly or monthly levels. The second view displayed how much the weekly activity has reduced the user's risks of a) overall mortality, b) dementia and c) cardiovascular health conditions (Figure 1b). ${ }^{8}$ The third view

[^3]

Figure 1: The information treatment view
Notes: The figure shows the new content included in the application for the treatment group. Figure 1(a) shows the default view when opening the application. The icons on the default view lead to views 1(b) and 1(c).
showed the user's rank in terms of total walking and cycling as compared to all application users (Figure 1c).

In addition to these elements, the treatment group also received an information message each Monday. The messages were sent as push notifications, which appeared irrespective of whether the application was opened. The short messages had varying content each week, which related to the overall recommendation for weekly activity, the user's personal activity level the past week or social comparisons to other application users' activity levels. The weekly messages are documented in Appendix C.

Our intervention combined pure information provision, including total amounts of walking and cycling, with health interpretations of this information, and social comparisons
including the user's rank compared to the full treatment group. All the information was personalized based on the individual-level data collected by the application. We chose to employ several channels for relaying information in order to increase the impact of the intervention. However, this decision meant that we were not be able to differentiate between the effectiveness of these different channels. This choice was driven by the limited sample size which prohibits splitting the treatment into groups based on the mechanism.

### 2.3 Randomization and summary statistics

We randomized application users in the Pirkanmaa region (5296 individuals) into two groups at the device level. Half of the sample was assigned to the treatment group with the new content in the application and the individualized health messages. The other half constitutes the control group, who continue to use the standard version of the carbon footprint calculator.

The distributions of the main variables of interest, walking and cycling length, are very skewed. The sample also includes a notable share of individuals who appear in the data at the beginning of data collection but are not actively using the application close to the start of treatment. We therefore created strata using the following variables:

1. pre-treatment levels (lengths) of walking, such that participants are split into five equally large groups based on their activity in August 2022 (20th percentiles) and classified as zero if not active,
2. pre-treatment levels of cycling in August 2022, such that participants are split into three groups based on their activity (33th percentiles) and zero if not active, and
3. geographical area, including three regions: Central Tampere ${ }^{9}$, suburban Tampere (other postal codes in Tampere) and other municipalities in Pirkanmaa. See Figure A. 1 for a visual representation of these regions.

Randomization was carried out in Stata using command randomize (seed 310). We ended up with 2649 individuals in the treatment group and 2647 individuals in the control group. The size of the groups differ by two individuals because of stratification. Table 1 presents summary statistics and shows the balance by assignment group.

[^4]Table 1: Summary statistics by treatment status

|  | Control | Treatment | P-value |
| :--- | :---: | :---: | :---: |
| Panel A: Postal code characteristics |  |  |  |
| Population density | 2251 | 2256 | 0.896 |
| Mean age | 41 | 41 | 0.512 |
| Median income | 22023 | 22077 | 0.371 |
| Pensioner share | 0.23 | 0.23 | 0.606 |
| Public transit stops per sq. km | 12 | 12 | 0.699 |
| City center locations, share | 0.22 | 0.24 | 0.157 |
| Panel B: travel behaviour |  |  |  |
| Api calls per day | 0.78 | 0.94 | 0.231 |
| Trips per day | 4.97 | 4.97 | 0.992 |
| Walking (km) | 1.67 | 1.64 | 0.524 |
| Cycling (km) | 0.91 | 1.03 | 0.243 |
| Bus (km) | 1.95 | 1.70 | 0.161 |
| Car (km) | 28.58 | 28.97 | 0.734 |
| Observations | 2641 | 2645 |  |

Notes: The table presents mean residential area characteristics (Panel A) and travel behaviour (Panel B) for the control and treatment group, and the p-value for the equality of means. Values calculated from the pretreatment period and standard errors for the means clustered at the device level. Application programming interface (API) calls measure application use.

Attrition We observe changes in the number of devices using the carbon footprint calculator throughout the observation period. The development is similar in the treatment and control group. During the treatment period, the number of devices observed at the weekly level declines steadily and similarly in both groups. The declining trend was present already before the treatment began (see Figure A. 3 in Appendix A). Based on this evidence, we conclude that attrition is not correlated with treatment status.

## 3 Empirical strategy

We estimate the average treatment effect using the following equation:

$$
\begin{equation*}
Y_{i t}=\beta T_{i}+\theta \bar{Y}_{i, P R E}+\alpha_{t}+\gamma^{\prime} \boldsymbol{X}_{\boldsymbol{i t}}+\varepsilon_{i t} \tag{1}
\end{equation*}
$$

Equation (1) is an ANCOVA regression. ${ }^{10}$ The indices $i$ and $t$ denote the user (device id.) and day using only post-treatment observations. $Y_{i t}$ is the outcome variable in levels or its inverse hyperbolic sine -transformation. Primary outcome variables include daily walking length, daily walking time and daily application use. Secondary outcome variables include daily cycling, driving, public transport use and the carbon footprint. The average treatment effect is measured by $\beta$, which is the coefficient on the treatment group indicator $T_{i} . \bar{Y}_{i, P R E}$ is the individual-level variable of interest averaged over the pre-experiment period. We use daylevel fixed-effects, $\alpha_{t}$. The covariates $\boldsymbol{X}_{\boldsymbol{i} \boldsymbol{t}}$ include characteristics that vary over individuals or individuals and days, such as location characteristics. We also include indicators for the randomisation strata in the list of covariates. Last, $\epsilon_{i t}$ is the error term.

We analyze heterogeneity in the treatment effect with respect to pre-treatment activity levels and local characteristics. We also explore heterogeneity in the time dimension, with respect to time-of-the-week and treatment duration. For the heterogeneity analysis, we estimate versions of (1) where we introduce a separate treatment group indicator for low and high levels in the above mentioned dimensions to capture heterogeneity. Specifically, the equation is the following:

$$
\begin{equation*}
Y_{i t s}=\beta_{s} T_{i} Z_{s}+\theta \bar{Y}_{i, P R E}+Z_{s}+\alpha_{t}+\gamma^{\prime} \boldsymbol{X}_{\boldsymbol{i} \boldsymbol{t}}+\varepsilon_{i t s} \tag{2}
\end{equation*}
$$

The indicator variable $Z_{s}$ groups the individuals into categories $s$ for the relevant measure of heterogeneity.

For the end-of-treatment survey we only have one cross-section of the data, and we are not able to introduce pre-treatment averages. To estimate the effect of treatment $T_{i}$ on survey answers, we estimate the following equation:

$$
\begin{equation*}
Y_{i}=\beta_{q} T_{i}+\gamma_{q}^{\prime} \boldsymbol{X}_{\boldsymbol{i}}+\varepsilon_{i}, \tag{3}
\end{equation*}
$$

where subindex $q$ refers to questionnaire. We use the respondents' self-reported age, gender and the number of children in the household as control variables. In all estimations, we cluster standard errors at the level of device ids.

[^5]

Figure 2: Average API calls by date and day of the week.
Notes: Graph (a) plots the daily application views for the control and treatment group. The vertical lines indicate the beginning and end of treatment, and the underlying grid lines indicate Mondays. Graph (b) shows mean application views and $95 \%$ confidence intervals by weekday in the control and treatment group during treatment. Standard errors for the means are clustered at the device level.

## 4 Results

We first examine whether the treatment group viewed the information that was transmitted to them in the form of the activity calculator and the weekly notification messages. We measure this pass-through of information using the number of daily application programming interface (API) calls for each device. ${ }^{11}$ Then, we turn to our survey responses to study whether the information is incorporated into the respondents' beliefs. Last, we analyze whether information impacted the observed travel behavior.

Is information received? Figure 2a displays the average number of daily API calls by treatment status. The vertical solid lines mark the beginning and end of treatment. The vertical grid lines indicate Mondays, the weekday when the treatment group received the weekly information messages. During the treatment period, application use is clearly higher in the treatment group. Furthermore, the spikes in use appear on Mondays, indicating that the notification messages were not ignored but instead incited application views. Such Monday spikes are not observed in the control group, where spikes are less pronounced and appear to occur randomly across weekdays.

Figure 2b provides a closer look at application use by weekday. The figure displays average application views by day and treatment status during the treatment period. The treatment group now clearly shows increased application use on Mondays. For the control group no specific weekday stands out, but there appears to be a declining interest in using the application towards the end of the week. On average, the treatment group engages with the application 1.30 times per day, and the control group 0.71 times per day.

We estimate the treatment effect on application use by estimating Equation (1) on individual-level daily data of application use. For estimating the intensive margin of application use, the dependent variable is the number of API calls sent by the application each day. For estimating the extensive margin of application use, the dependent variable is a daily indicator that takes value one if at least one API call is sent by the application. Table B. 1 presents the results. Intensive margin results are presented in columns 1 and 2, extensive margin results in columns 3 and 4. The average treatment effect on the number of application views is clearly positive and highly statistically significant. Conditional on

[^6]pre-treatment levels of application use, the treatment group views the application on average 0.36 times more per day, which corresponds to a 51 percent increase compared to the average value in the control group. Column 2 splits the treatment effect by treatment month. Interestingly, the treatment effect does not seem to fade out as time proceeds, as indicated by the coefficients on treatment months.

These results show that the information is received by the treatment group. However, the results on the extensive margin of application use show no differences between the treatment and control group. This indicates that the treatment did not incentivize inactive application users to become more engaged. The effect is thus coming from more intensive use by active application users.

Is information internalized? We aim to understand whether the respondents internalized the information that they received. We explore whether treatment and control groups respond differently to the end-of-treatment survey, which measures their knowledge about health benefits and general activity levels in the population. The treatment effect is shown in Figure 3.

The first panel (Fig. 3a) shows the knowledge question "According to your assessment, what share of Finns aged over 30 commute to work by cycling or walking?". The treatment group did not receive a direct answer on this particular question, but they were given information on the activity by overall application users (see Figure 1c). We find no statistically significant differences between the control and treatment groups.

The second panel (Fig. 3b) shows the other knowledge question: "According to your assessment, how much does regular, sufficient activity reduce the overall risk of mortality?. This was the personalized health information given to individuals (see Figure 1b). We find that the treatment increases the likelihood of knowing the correct answer. This suggests that the health information has been internalized by the respondents. Table B. 2 in Appendix A shows additional analyses, which reveal that the control group does not systematically under- or overestimate the health benefits. The information therefore reduced the variance of answers, rather than shifting the mean on the scale.


Figure 3: Survey results: Is information internalized?
Notes: Graph (a) presents the proportion of correct answers by treatment status to the question "According to your assessment, what share of Finns aged over 30 commute to work by cycling and walking?". Graph (b) presents the proportion of correct answers by treatment status to the question "According to your assessment, how much does regular, sufficient activity reduce the overall risk of mortality?". $95 \%$ confidence intervals are illustrated by the vertical lines, standard errors for the means are clustered at the device level.

Does information change travel behavior? Our main variable of interest is the amount of walking. ${ }^{12}$ Figure 4 a shows the average daily walking length for the treatment and control groups. Again, the vertical lines indicate the beginning and end of treatment and the vertical gridlines indicate Mondays. No clear difference in levels between the groups is visible during the treatment period. Also, while the spikes indicate a weekly pattern in walking, a single day does not stand out as clearly as in the application use data.

Figure 4b displays average walking amounts by weekday and treatment status during the treatment period. The weekly pattern is now visible: in both groups, walking increases during the weekend. Overall, the daily levels of walking are similar across the groups.

[^7]

Figure 4: Average walking by date and day of the week.
Notes: Graph (a) plots the daily levels of walking for the control and treatment group. The vertical lines indicate the beginning and end of treatment, and the underlying grid lines indicate Mondays. Graph (b) shows mean walking and $95 \%$ confidence intervals by weekday in the control and treatment group during treatment. Standard errors for the means are clustered at the device level.

In Appendix B, Table B. 3 presents results from estimating Equation (1) on daily observations of walking length and time during the treatment period. The estimations confirm the visual evidence in Figure 4: we do not find a statistically significant effect of treatment on either variable of interest. The estimation result in levels indicates a treatment effect of 27 meters per day, corresponding to $2 \%$ of the mean level of walking in the control group during treatment ( 1570 meters). We can rule out treatment effects higher than 110 meters per day at the $95 \%$ confidence level.

Although the average treatment effect is indistinguishable from zero, certain groups might still respond to the information. We examine heterogeneity in the treatment effect with respect to (i) time of week (weekday/weekend), (ii) the pre-treatment level of walking (splitting at the median value of daily walking), (iii) availability of public transit (below/above median value of transit stops per kilometer at home location), (iv) residential area income level (splitting at the median of postal code median income) and (v) intensity of application use (splitting at those who use the app on average once per day versus less frequent use). ${ }^{13}$ We do not find statistically significant treatment effects in any of these groups, nor are the differences between them statistically significant. The null results we find are in that sense robust.

The secondary variables of interest include the daily amount of driving, cycling, use of public transit and the daily carbon footprint calculated by the application. We estimate equation (1) in levels and find no statistically significant effects of the treatment on these variables. ${ }^{14}$ Figure 5 displays results from these estimations with the treatment effect scaled by the mean level of the dependent variable in the control group during treatment. The figure illustrates how estimation results have expected signs and magnitudes across the transport modes are reasonable. However, we do not have enough statistical power to detect the potentially very small average effects of the treatment. Note that cycling is not shown in the figure, to enhance the visibility of the effects on other transport modes. Due to very noisy data, the estimated effect on cycling is very imprecise and is of a different magnitude than the other effects.

In addition to estimation in levels, we have examined mode shares in terms of daily or weekly travel length, as well as the extensive margin of using a given transport mode conditional on observing any travel on that day. The results on mode shares and the probability

[^8]

Figure 5: Average treatment effect, proportional to control group mean.
Notes: The figure displays the average treatment effect and $95 \%$ confidence intervals on the daily length of different transport modes, expressed in proportion to the daily mean in the control group during treatment. The treatment effect is estimated according to Equation (1), standard errors clustered at the level of devices.
of using a given transport mode mirror those reported previously: we find no statistically significant treatment effects. Results are reported in Appendix B in Table B.8.

The role of carbon footprint information The information treatment takes place within a carbon footprint calculator, which is designed to give personalized information on the environmental impact of transport choices. The goal of the calculator is to help the city of Tampere reach its climate goals, which include increasing the modal share of public transport. If the carbon footprint information is successful in changing transport choices, it could be that individuals have already adapted their behaviour, and this is why our additional information provision does not bring about further changes. If this was the case, we would expect to see a negative association between the carbon footprint and the duration of application use. However, we find no correlation between these variables, neither in the sample overall nor when accounting for seasonality or the availability of public transport.

## 5 Conclusions

The negative health effects of low physical activity can be mitigated through governmental policies that discourage car use or promote active modes of transportation, thereby substantially reducing the personal costs. But are internalities a reason for government intervention? We study a specific internality: one stemming from insufficient information related to the benefits of physical activity, the recommendations regarding activity, or the activity of peers.

We design an information treatment where individuals are found to receive information about the health benefits of activity, and update their beliefs accordingly. However, this information does not lead to statistically significant changes in observed behavior, and we can rule out effects that are economically meaningful. Our results show that updating on the information leads to the same travel choices as before, therefore giving some evidence for the "rational choice" model of inactivity.

However, other behavioral biases, such as commitment problems (see Feil et al. 2023; Gruber and Kőszegi 2004; Della Vigna and Malmendier 2006) or habit formation, may justify policy intervention. Future research could provide further insights into these biases. Furthermore, especially in the context of transport, individuals' choices can be constrained by long-term fixed choices such as the type of residential area. Removing such constraints can entail large costs, which makes it difficult to act on newly acquired information in the short run.

Our results therefore do not exhaustively rule out the need for interventions. Yet, it appears that transport behaviour is difficult to change through information provision, even in the case of highly personalized and individually relevant information such as our treatment provided. Despite this, we are carefully optimistic about the potential of mobility tracking applications, because personalized information can provide value to the users and produce data for researchers at a relatively low cost.

## References

Bennett, G.G., Wolin, K.Y., Puleo, E.M., Mâsse, L.C., Atienza, A.A., 2009. Awareness of national physical activity recommendations for health promotion among us adults. Medicine and science in sports and exercise 41, 1849.
van den Bijgaart, I., Klenert, D., Mattauch, L., Sulikova, S., 2020. Healthy climate, healthy bodies: Optimal fuel taxation and physical activity. Economica .

Cellina, F., Bucher, D., Mangili, F., Veiga Simão, J., Rudel, R., Raubal, M., 2019. A large scale, app-based behaviour change experiment persuading sustainable mobility patterns: Methods, results and lessons learnt. Sustainability 11. URL: https://www.mdpi.com/ 2071-1050/11/9/2674, doi:10.3390/su11092674.

Costa Santos, A., Willumsen, J., Meheus, F., Ilbaw, A., Bull, F.C., 2022. The cost of inaction on physical inactivity to healthcare systems. Available at SSRN 4248284.

Coups, E.J., Hay, J., Ford, J.S., 2008. Awareness of the role of physical activity in colon cancer prevention. Patient education and counseling 72, 246-251.

Della Vigna, S., Malmendier, U., 2006. Paying not to go to the gym. american economic Review 96, 694-719.

Ding, D., Lawson, K.D., Kolbe-Alexander, T.L., Finkelstein, E.A., Katzmarzyk, P.T., Van Mechelen, W., Pratt, M., Committee, L.P.A.S..E., et al., 2016. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. The Lancet 388, 1311-1324.

Feil, K., Fritsch, J., Rhodes, R.E., 2023. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. British Journal of Sports Medicine .

Garcia, L., Pearce, M., Abbas, A., Mok, A., Strain, T., Ali, S., Crippa, A., Dempsey, P.C., Golubic, R., Kelly, P., et al., 2022. Non-occupational physical activity and risk of 22 cardiovascular disease, cancer, and mortality outcomes: a dose-response meta-analysis of large prospective studies. medRxiv, 2022-03.

Gessner, J., Habla, W., Wagner, U., 2023. Can social comparisons and moral appeals increase public transport ridership and decrease car use? ZEW - Centre for European Economic Research Discussion Paper doi:http://dx.doi.org/10.2139/ssrn. 4350517.

Goetz, A., Marinica, I., Mayr, H., Mosetti, L., Schubert, R., 2023. Do mobile applications foster sustainable mobility? evidence from a field experiment. Available at SSRN doi:http://dx.doi.org/10.2139/ssrn. 4310386.

Golman, R., Hagmann, D., Loewenstein, G., 2017. Information avoidance. Journal of economic literature 55, 96-135.

Gosnell, G., Martin, R., Muûls, M., Coutellier, Q., Strbac, G., Sun, M., Tindermans, S., 2019. Making smart meters smarter the smart way. CEP Discussion Paper No 1602 URL: https://eprints.lse.ac.uk/102618/1/dp1602.pdf.

Gravert, C., Olsson Collentine, L., 2021. When nudges aren't enough: Norms, incentives and habit formation in public transport usage. Journal of Economic Behavior \& Organization 190, 1-14. URL: https://www.sciencedirect.com/science/article/pii/S0167268121003000, doi:https://doi.org/10.1016/j.jebo.2021.07.012.

Gruber, J., Kőszegi, B., 2004. Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. Journal of Public Economics 88, 1959-1987.

Haase, A., Steptoe, A., Sallis, J.F., Wardle, J., 2004. Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. Preventive medicine 39, 182-190.

Hintermann, B., Molloy, J., Beaumont, S., Götschi, T., Castro, A., Tchervenkov, C., Tomic, U., Axhausen, K., 2022. Pigovian transport pricing in practice. WWZ Working Paper URL: https://edoc.unibas.ch/88604/.

Kohl, H.W., Craig, C.L., Lambert, E.V., Inoue, S., Alkandari, J.R., Leetongin, G., Kahlmeier, S., 2012. The pandemic of physical inactivity: global action for public health. The lancet 380, 294-305.

Kristal, A., Whillans, A., 2020. What we can learn from five naturalistic field experiments that failed to shift commuter behaviour. Nature Human Behaviour 4, 169-176. URL: https://doi.org/10.1038/s41562-019-0795-z.

Lieberoth, A., Holm Jensen, N., Bredahl, T., 2018. Selective psychological effects of nudging, gamification and rational information in converting commuters from cars to buses: A controlled field experiment. Transportation Research Part F: Traffic Psychology and Behaviour 55, 246-261. URL: https://www.sciencedirect.com/science/article/pii/ S1369847817304527, doi:https://doi.org/10.1016/j.trf.2018.02.016.

Löschel, A., Rodemeier, M., Werthschulte, M., 2023. Can self-set goals encourage resource conservation? field experimental evidence from a smartphone app. European Economic Review 160, 104612. doi:https://doi.org/10.1016/j.euroecorev.2023.104612.

McKenzie, D., 2012. Beyond baseline and follow-up: The case for more t in experiments. Journal of development Economics 99, 210-221.

Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F.B., Golubic, R., Kelly, P., Khan, S., Utukuri, M., Laird, Y., et al., 2022. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA psychiatry .

Publishing, O., for Economic Co-operation, O., Development, 2012. Mortality risk valuation in environment, health and transport policies. OECD.

Schwab, L., Hintermann, B., Goetz, A., Molloy, J., 2023. Reducing transportation externalities through nudging: Results from a gps-tracked experiment. Unpublished manuscript, available from authors by request .

## Appendices

## A Data

Figure A. 1 shows Pirkanmaa region in Finland, which is our study area. Black lines indicate postal code boundaries, and colors indicate the three regions used in stratification. Red is central Tampere ( postal codes 33230, 33200, 33210, 33180, 33500, 33540 and 33100), Yellow is suburban Tampere (other postal codes in Tampere) and blue are postal codes in other municipalities in Pirkanmaa.


Figure A.1: Study area
Notes. The map shows the study area, and the geographical stratification in three layers: central Tampere (red), other Tampere (yellow) and other postal codes in Pirkanmaa (blue). The lines indicate postal code boundaries.

Figure A. 2 illustrates the representativeness of the data. Each dot represents one postal code, with the x -axis showing the application users and the y -axis the total number of inhabitants in that postal code area (in logs). The figure shows that our data include users from areas of different sizes, and not only from highly populated areas such as central Tampere.

Table A. 1 compares our sample to the results of the National Travel Survey (HLT) from


Figure A.2: Geographical representativeness of the sample
Notes. The figure plots the number of users per postal code (x-axis) and the total number of inhabitants in the same postal code. The blue line shows the fit (inhabitants $=6.10+0.63 \times$ users). Logarithmic transportation is used to enhance readability of the graph.

2016 and 2021. The table includes results for the subsample representing the Pirkanmaa region. Our sample includes somewhat more walking and less public transport use in terms of trip lengths. Our sample has significantly longer walking times, because the application gives longer duration for walking, as the application counts the entire trip duration to its length (e.g. a visit to a store or a playground). Our data do not include the category "other".

Figure A. 3 illustrates the changes to the number of unique devices observed each week by treatment status. The vertical line indicates the beginning of treatment. A declining trend is visible already before the beginning of treatment, and this trend continues throughout the treatment period similarly in both groups.

Table A.1: Representativeness of the sample compared to earlier surveys (National Travel Survey (HLT) from 2016 and 2020)

|  | $\begin{gathered} \text { Length } \\ \text { (km/day) } \end{gathered}$ |  |  | $\begin{gathered} \text { Time } \\ \text { (minutes/day) } \end{gathered}$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | This study | HLT 2016 | HLT 2021 | This study | HLT 2016 | HLT 2021 |
| Walking | 1.6 | 1.2 | 1.2 | 43.4 | 18 | 18 |
| Biking | 0.9 | 0.8 | 0.7 | 3.9 | 4 | 4 |
| Bus | 1.9 | 3.9 | 2.1 | 4.6 | 10 | 6 |
| Tram and train | 1.2 | 2.9 | 2.1 | 1.3 | 2 | 2 |
| Car | 28.1 | 30 | 28.9 | 34.8 | 39 | 38 |
| Other | - | 1.2 | 0.7 | - | 3 | 2 |
| N | 5288 | 4183 | 2859 | 5288 | 4183 | 2859 |

Notes: The table presents a comparison of our sample to two most recent waves of the National Travel Survey's subsample on the region of Pirkanmaa. The results from National Travel Surveys are reported by the Finnish Transport and Communications Agency and can be found at https://www.traficom.fi/fi/hlt.


Figure A.3: Number of unique devices observed each week by treatment status
Notes. The figure plots the number of unique devices observed weekly in the control and treatment groups. The vertical line indicates the beginning of the treatment period.

## B Results

## B. 1 Application use

Table B. 1 presents results from estimating the ANCOVA Equation (1). The dependent variable is either application views per day (API calls) or an indicator for observing at least one API call per day.

Table B.1: The effect of treatment on application use.

| Dependent variable: | Intensive margin Number of app views |  | Extensive margin |  |
| :---: | :---: | :---: | :---: | :---: |
|  | (1) | (2) | (3) | (4) |
| Treatment | 0.359*** |  | 0.005 |  |
|  | (0.010) |  | (0.011) |  |
| Treatment by month |  |  |  |  |
| September |  | 0.331*** |  | 0.005 |
|  |  | (0.104) |  | (0.010) |
| October |  | 0.468*** |  | 0.009 |
|  |  | (0.139) |  | (0.013) |
| November |  | 0.243** |  | -0.001 |
|  |  | (0.109) |  | (0.015) |
| Pre-treatment level | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Daily FE | Yes | Yes | Yes | Yes |
| Mean in control group | 0.71 | 0.71 | 0.29 | 0.29 |
| Observations | 44659 | 44659 | 44659 | 44659 |
| Devices | 1012 | 1012 | 1010 | 1012 |

Notes: The table displays results from estimating the average treament effect on application use, Equation (1). Covariates include stratification groups (level of walking, level of cycling, location cateogory, and residential location characteristics (public transit stops per square km, population density, share of pensioners, median income). Standard errors clustered at device level.

## B. 2 Survey results

Table B. 2 shows the estimated treatment effect on two survey questions, as indicated in the two panels. Columns 1-2 show the results for the levels (in Likert scale) and columns 3-4 show the result as shares of correct answers. The first questions asks "According to your assessment, how much does regular, sufficient activity reduce the overall risk of mortality? ( $0-10 \%, 10-20 \%, 20-30 \%, 30-40 \%, 40-50 \%$ ) and, second, "According to your assessment, what share of Finns aged over 30 commute to work by cycling or walking? " ( $0-20 \%, 20-40 \%$, $40-60 \%, 60-80 \%, 80-100 \%)$.

Table B.2: Survey results: is information internalized?

| Answer on Likert scale |  | Indicator for right answer |  |
| :---: | :---: | :---: | :---: |
| $(1)$ | $(2)$ | $(3)$ | $(4)$ |

Question A: How much does sufficient activity reduce the mortality risk?

| Treatment | 0.033 | 0.084 | $0.149^{* *}$ | $0.125^{*}$ |
| :---: | :---: | :---: | :---: | :---: |
|  | $(0.152)$ | $(0.151)$ | $(0.061)$ | $(0.064)$ |

Question B: What share of Finns commute to work by cycling or walking?

| Treatment | -0.062 | -0.101 | -0.029 | -0.043 |
| :--- | :---: | :---: | :---: | :---: |
|  | $(0.117)$ | $(0.127)$ | $(0.084)$ | $(0.091)$ |
|  |  |  |  |  |
| Covariates |  | Yes |  | Yes |
| Individuals | 141 | 141 | 141 | 141 |

Notes: The table presents results from estimating Equation (3) on answers to the end-of-treatment survey (OLS estimation). Columns 1-2 show results for the answers on a Likert scale (range 1 to 5). Columns 3-4 show results for an indicator variable that takes value one for a correct answer and zero otherwise. Covariates include self-reported age, gender, and presence of underage children in the households.

## B. 3 Walking

Table B. 3 presents results from estimating Equation (1) with daily walking as the dependent variable. Walking is measured in meters per day or minutes per day or the asinhtransformation of these variables.

Table B.3: Effect of treatment on walking

|  | Walking length (km) |  | Walking time (min) |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Linear | Asinh | Linear | Asinh |
|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ |
| Treatment | 26.824 | -0.090 | 0.338 | -0.024 |
|  | $(42.857)$ | $(0.085)$ | $(0.958)$ | 0.027 |
|  |  |  |  |  |
| Pre-treatment level | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Daily FE | Yes | Yes | Yes | Yes |
|  |  |  |  |  |
| Mean in control group | 1570 |  | 42 |  |
| Observations | 46506 | 46506 | 46506 | 46506 |
| Devices | 1309 | 1309 | 1309 | 1309 |

Notes: The table displays results from estimating the average treatment effect on daily walking, Equation (1). Covariates include stratification groups (level of walking, level of cycling, location cateogory), and residential location characteristics (public transit stops per square km, population density, share of pensioners, median income). Standard errors clustered at device level.

Table B. 4 presents results from estimating Equation (2) on daily levels of walking measured in meters.

Table B.4: Walking: heterogeneity in treatment effect

|  | Application <br> use intensity | Pre-treatment <br> walking level | Income in <br> residential location | Public transit <br> availability | Weekend |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Low $\times$ Treatment | 8.763 | 10.288 | -9.815 | 15.721 |  |
|  | $(50.570)$ | $(50.596)$ | $(68.155)$ | $(61.558)$ |  |
| High $\times$ Treatment | 71.225 | 48.357 | 49.872 | 41.269 |  |
| Weekday $\times$ Treatment | $(82.811)$ | $(71.499)$ | $(55.435)$ | $(59.991)$ | 24.941 |
|  |  |  |  | $(45.794)$ |  |
| Weekend $\times$ Treatment |  |  |  | 32.167 |  |
|  |  |  |  | $(60.223)$ |  |
| Pre-treatment level | Yes | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes |
| Daily FE | Yes | Yes | Yes |  | Yes |
| Observations | 46506 | 46506 | 46506 | 46506 | 46506 |
| Devices | 1309 | 1309 | 1309 | 1309 | 1309 |

Notes: The table displays heterogeneity in the average treatment effect by application use intensity, pre-treatment walking level, income in the residential location, public transit availability and time of week. The dependent variable in all estimations is the level of walking, measured in meters per day. Covariates include stratification groups (level of walking, level of cycling, location cateogory), and residential location characteristics (public transit stops per square km, population density, share of pensioners, median income). Standard errors clustered at device level.

## B. 4 Secondary outcome variables

Table B. 5 presents results from estimating Equation (1) with the daily level of driving, public transport use or cycling as the dependent variable. The level is measured in kilometers per day. The carbon footprint refers to the amount of $\mathrm{CO}_{2}$ emitted from daily transport, measured in grams.

Table B.5: Effect of treatment on travel outcomes

|  | Driving <br> (km per day) | Public <br> (km per day) | Cycling <br> (km per day) | $\mathrm{CO}_{2}$ <br> $\left(\mathrm{gCO}_{2}\right.$ per day) |
| :--- | :---: | :---: | :---: | :---: |
| Treatment | -0.276 | 0.032 | -0.161 | -87.488 |
|  | $(0.939)$ | $(0.143)$ | $(0.109)$ | $(172.145)$ |
| Pre-treatment level | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Daily FE | Yes | Yes | Yes | Yes |
| Mean in control group | 26 | 0.29 | 0.93 | 4876 |
| Observations | 46533 | 46533 | 46533 | 46533 |
| Devices | 1309 | 1309 | 1309 | 1309 |

Notes: The table displays results from estimating the average treatment effect (Equation (1)) with daily driving, public transit use, cycling or carbon footprint as the dependent variable. The transport modes are measured in kilometers per day and the carbon footprint in grams of $\mathrm{CO}_{2}$ per day. Covariates include stratification groups (level of walking, level of cycling, location cateogory), and residential location characteristics (public transit stops per square km, population density, share of pensioners, median income). Standard errors clustered at device level.

## B. 5 Additional analyses

Extensive margin: probability of mode use The following extensive margin calculations are based on observing car or public transit use, walking or cycling on a day for which at least one transport mode is observed. We do not include the days for which no observation of the device is available, as this would require making assumptions on why the application is not tracking.

Table B. 6 presents mean shares of daily mode use during treatment as well as correlations between daily mode use. Average daily mode use reveals that walking is constant across the week whereas public transit use is more likely on weekdays and car use increases during weekends.

Correlations are presented for weekdays, to represent mode use in commuting and errands. Observing walking is positively correlated with observing public transit use, and negatively correlated with car use. Analogously, there is a negative correlation between public transit use and car use.

Table B.6: Extensive margin of transport mode use: means and correlations

|  | Walking | Public transit use | Car use | Cycling |
| :--- | :---: | :---: | :---: | :---: |
| Panel A: Mean daily mode use |  |  |  |  |
| Weekday | 0.81 | 0.24 | 0.64 | 0.17 |
| Weekend | 0.80 | 0.15 | 0.69 | 0.13 |
|  |  |  |  |  |
| Correlations during weekdays | 0.17 | -0.16 | -0.06 |  |
| Walking | 1.00 | 1.00 | -0.23 | -0.08 |
| Public transit use |  | 1.00 | -0.04 |  |
| Car use |  |  |  |  |
| Observations: 46538 |  |  |  |  |

Notes: The table presents mean daily mode use, conditional on observing any travel on that day (Panel A). Panel B shows correlations between observed daily mode use. Observations are from the treatment period.

Table B. 7 presents the average treatment effect on the extensive margin of observing walking, public transit use, car use and cycling. The dependent variable is a binary indicator taking value one if the transport mode is observed on a given day. The estimated effect is zero for other modes than cycling, where we observe that the treatment group reduced their probability of cycling during the treatment.

Table B.7: Effect of treatment on probability of daily mode use

|  | Walking | Public transit use | Car use | Cycling |
| :--- | :---: | :---: | :---: | :---: |
| Treatment | -0.007 | 0.000 | -0.002 | $-0.025^{* * *}$ |
|  | $(0.008)$ | $(0.009)$ | $(0.009)$ | $(0.009)$ |
|  |  |  |  |  |
| Pre-treatment level | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes |
| Daily FE | Yes | Yes | Yes | Yes |
| Mean in control group | 0.81 | 0.21 | 0.66 | 0.18 |
|  |  |  |  |  |
| Observations | 46533 | 46533 | 46533 | 46533 |
| Devices | 1309 | 1309 | 1309 | 1309 |

Notes: The table displays results from estimating the average treatment effect (Equation (1)) on the daily probability of using each transport mode. The dependent variable is an indicator taking value one if the transport mode is observed, conditional on having observations of any travel on that day. Covariates include stratification groups (level of walking, level of cycling, location cateogory), and residential location characteristics (public transit stops per square km, population density, share of pensioners, median income). Standard errors clustered at device level.

Mode shares and mode substitution The intervention aims to increase active transport. There are three options for doing this:

1. The treatment group switches to using public transit instead of the car. We would then observe no difference in trip lengths, because destinations don't change. The share of public transit and walking would increase. The total length of walking and public transit would also increase, as long as walking for leisure is not reduced due to walking more during daily transport.
2. The treatment group changes destinations so as to avoid car use and uses active transport or public transit instead. In this case total trip length would go down. Again, the share of walking and public transit would increase. The total lengths of active transport and/or public transit would also increase, as long as other walking is not reduced.
3. The treatment group increases walking by going for leisure walks but makes no changes to commuting or errands. In this case, total trip length would increase and the total amount of walking would increase. The share of walking would also increase.

Table B.8: Effect of treatment on daily and weekly mode shares.

|  | Walking | Cycling | Public transit | Car |
| :--- | :---: | :---: | :---: | :---: |
| Panel A: Dependent variable: daily mode share |  |  |  |  |
| Treatment | 0.003 | $-0.012^{*}$ | 0.004 | 0.005 |
|  | $(0.007)$ | $(0.006)$ | $(0.007)$ | $(0.008)$ |
| Mean in control group during treatment | 0.22 | 0.08 | 0.18 | 0.52 |
| Observations | 46506 | 46506 | 46506 | 46506 |
| Individuals | 1309 | 1309 | 1309 | 1309 |
| Panel B: Dependent variable: weekly mode share |  |  |  |  |
|  |  |  |  |  |
| Treatment | 0.003 | -0.003 | 0.002 | 0.003 |
|  | $(0.006)$ | $(0.005)$ | $(0.009)$ | $(0.007)$ |
| Mean in control group during treatment | 0.14 | 0.04 | 0.19 | 0.60 |
| Observations | 9526 | 9526 | 9526 | 9526 |
| Individuals | 1309 | 1309 | 1309 | 1309 |

Notes: The table displays results from estimating the average treatment effect (Equation (1) with daily or weekly mode share as the dependent variable. The mode share is calculated as daily or weekly mode length in proportion to total daily or weekly trip length. Covariates include stratification groups (level of walking, level of cycling, location category), and residential location characteristics (public transit stops per square km , population density, share of pensioners, median income). Standard errors clustered at device level.

It is thus possible that substitution between mode shares is happening even if we do not observe increases in total walking length. If this is the case, there would be no health benefits but the carbon footprint would decrease. The mode substitution would show up in mode shares.

Table B. 8 reports results from estimating Equation (1) with the daily or weekly mode share as the dependent variable. The estimated effects are very small in magnitude and for the most part not statistically significant. We have analysed heterogeneity with respect to application use, the availability of public transit and the treatment duration, and find no statistically significant effects or differences between the groups. We also find no changes to total trip lengths on a daily or weekly level.

## C Details of the design: weekly information messages






[^0]:    *Lassi Ahlvik (lassi.ahlvik@helsinki.fi): University of Helsinki, Department of Economics and Management. Anna Sahari (anna.sahari@vatt.fi): VATT Institute for Economic Research. We are grateful for funding from the Finnish Ministry of the Environment for project "Keli - Kestävämmän liikkumisen kehittäminen hiilijalanjälkilaskurin avulla" and from the Finnish Center of Excellence in Tax Systems Research funded by the Research Council of Finland (project 346253). We thank Anna Vilhula, Juha Yrjölä, Karlo Tuominen, Theo Blauberg, Juuso Koponen and Marko Tainio for the invaluable assistance in designing the experiment. The experiment was approved by The University of Helsinki Ethical Review Board in Humanities and Social Sciences (statement 9/2022) and pre-registered at the AEA RCT Registry (see RCT ID AEARCTR-0009600).

[^1]:    ${ }^{1}$ Ding et al. (2016) estimate that physical inactivity was responsible for the loss of 13.4 million disabilityadjusted life-years (DALYs) worldwide. For example, using a conservative estimate of $\$ 50,000$ Euros/DALY (see Publishing et al. 2012), the cost of inactivity from increased mortality is $\$ 536 \mathrm{bn} /$ year. The same study estimates the cost of physical inactivity on the healthcare system to be $\$ 54 \mathrm{bn}$ per year. Of this figure, $58 \%$ is paid by the public sector, $24 \%$ by the private sector and $18 \%$ by households. In addition, there are estimated $\$ 14 b n$ in productivity losses. In an updated analysis, Costa Santos et al. (2022) find a higher figure for the health care system costs, $\$ 47.6$ bn.
    ${ }^{2}$ As the social benefits of physical activity exceed the private benefits, the textbook response would be to align the two by imposing a subsidy on active travels. This first-best policy cannot be implemented, because it is impossible to observe and verify individuals' walking and cycling behavior.

[^2]:    ${ }^{3}$ Smartphone apps have also been used in the context of energy use by Löschel et al. (2023) to carry out a RCT related to energy conservation and by Gosnell et al. (2019) to provide feedback on electricity.
    ${ }^{4}$ The algorithm was developed in collaboration with Cognitive Science unit at University of Helsinki and a private company Kausal. The fine-grained detection of transportation modes used Interacting Multiple Model Kalman filter, which integrated (1) accelerometer-based mode detection, (2) real-time GPS information and (3) public transportation timetables.
    ${ }^{5}$ Figure A. 1 presents a map of our study area

[^3]:    ${ }^{6}$ To comply with data regulation, the treatment and control group received a different link, allowing us to track the assignment status of the respondent. However, we are not able to connect responses to device ids.
    ${ }^{7}$ The analysis was first pre-registered on June 17, 2022 and updated on August 31, 2022 before the treatment began, see RCT ID AEARCTR-0009600. We received a positive ethical statement (9/2022) from the University of Helsinki Ethical Review Board in Humanities and Social and Behavioral Sciences.
    ${ }^{8}$ These health outcomes are calculated using openly available health risk calculators provided by the University of Cambridge and available at https://shiny.mrc-epid.cam.ac.uk/ meta-analyses-physical-activity

[^4]:    ${ }^{9}$ Central Tampere included postal codes $33230,33200,33210,33180,33500,33540$ and 33100

[^5]:    ${ }^{10}$ For efficiency gains of ANCOVA compared to difference-in-differences estimator, see McKenzie (2012)

[^6]:    ${ }^{11}$ The application sends an API call when the application is opened, requests data or performs an action, allowing to track the users' engagement with the application.

[^7]:    ${ }^{12}$ In addition to being a mode of transport in itself, walking is also positively correlated with public transport use (see Appendix B, Table B.6). The measure of walking therefore represents both potential modal shifts.

[^8]:    ${ }^{13}$ Estimation results are available in Appendix B in Table B.4.
    ${ }^{14}$ Estimation results are documented in Appendix B Table B. 5 .

