Pre-analysis plan for study "Norm-based feedback on household waste: a large-scale field experiment in two Swedish municipalities"

1. Introduction

Over the last decade or so, there has been a marked upsurge of research into 'norm-based feedback' interventions, which aim to reduce the consumption of some resource by providing users with information about their use compared to similar peers, e.g. neighbors in their community. Many such studies have been carried out on energy or electricity use (Schultz et at, 2007; Allcott, 2011; Ayres et al, 2012; Costa & Kahn, 2013; Allcott & Rogers, 2014; LaRiviere et al, 2014; Allcott, 2015; Dolan & Metcalfe, 2015; Alberts et al, 2016; Andor et al, 2018; Allcott & Kessler, 2019) and water use (Ferraro & Price, 2013; Jaime Torres & Carlsson, 2018). However, as far as we are aware, no published study exists on the effect of providing household-specific norm-based feedback within the waste domain, which is the central contribution of our project.

We run separate randomized controlled trials of norm-based feedback on household residual waste (the unsorted waste fraction) in two municipalities in western Sweden, Partille and Varberg. In total, some 20,000 single-family houses participate in these studies, about three-quarters of which are located in Varberg. As in previous studies, the aim of the experiment is to examine whether treated households that receive feedback reduce their generation of residual waste compared to a control group that does not receive feedback.

Our main feedback condition, which we term "static" feedback, is a replication of the by now standard Opower-style norm (e.g. Ayres et al 2012; Allcott & Rogers, 2014). This involves mailing letters comparing each household's own amount of generated residual waste (in kilograms per person) with (i) average residual-waste generation among a set of roughly 100 "comparable" households, and with (ii) average waste generation among the "top" 20% within that set, i.e. the (roughly) 20 households that generated the least waste. We also explore differently presented, "dynamic" feedback. These letters emphasize changes from one period to the next and is inspired by Sparkman & Walton (2017), who provide evidence that norms focusing on *changes* in behavior, rather than snapshot pictures of consumption differences within a population, are more effective at reducing use.

Both participating municipalities had pay-as-you-throw incentives in place prior to the study period and weigh all waste bins during collection. We therefore collaborate with each municipal waste department to base all types of feedback on actual household-level data, so that the information provided is accurate. The feedback letters are stamped with the municipality logo and sent out every four or twelve weeks (depending on the treatment arm) across March-October 2019. At the bottom of each letter, we include an URL to a home page where participants can opt out of the experiment.

The structure of this document is as follows. Section 2 describes the parameters of the experiment, including feedback design, in greater detail. Section 3 lists data sources used by the project, and section 4 outlines our randomization methodology. Section 5 provides information on how the feedback is constructed from waste data, as well as other details regarding the intervention itself. Relatedly, section 6 outlines our procedures for constructing e.g. our main outcome variable. Section 7 presents an ex-ante power calculation. Finally, in section 8, we describe what statistical methods we will use to estimate treatment effects as well as perform various robustness tests and supplementary analyses.

2. Experimental design

The project includes two separate studies carried out in parallel during March-October 2019 in Partille and Varberg, two municipalities in western Sweden. As discussed in more detail in section 4 below, we use cluster randomization with blocking in both municipalities. We randomize in geographically contiguous clusters of households in order to mitigate potential interference between treatment and control households, which we suspect operates primarily between immediate neighbors. While there have been no major issues with such across-household spillovers when presenting norm-based feedback on energy or water use, we take the conservative approach in this project because we are applying norm-based feedback in a new domain.

In both municipalities, households are divided roughly equally into three treatment arms, including a control group. However, the two treatments differ across our participant municipalities, as shown in Table 1. In the Partille study, we vary the frequency of feedback, with one treatment group receiving feedback every four weeks ("monthly"), and the other receiving feedback every twelve weeks ("quarterly"). Households in the "monthly" condition receive a total of nine feedback letters, while households in the "quarterly" condition receive three feedback letters (see also sections 5 and 6). In the Varberg study, we explore a different type of feedback that focuses on the dynamic aspects of waste reduction. This is to be contrasted with the "static" norm feedback provided to all other treated households.

Partille	1. Control: no letters sent 2. Monthly norm-based feedback letter: "static" 3. Quarterly norm-based feedback: "static"
Varberg	1. Control: no letters sent 2. Monthly norm-based feedback letter: "static" 3. Monthly norm-based feedback letter: "dynamic"

Table 1. Experimental treatment arms in the two municipalities.

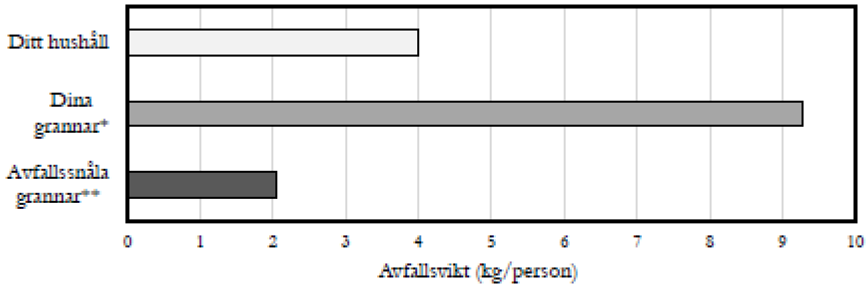
Figure 1 provides an example of static feedback provided to households in Varberg. The setup is very similar to previous studies on norm-based feedback such as Allcott & Rogers (2014) or Andor et al (2017). The bar chart on the upper part of the page displays, from top to bottom and for the latest comparison period (four weeks in the "monthly" condition, twelve weeks in the "quarterly" condition):

VIVAB strävar efter att minska det brännbara avfallet och vill med detta utskick informera dig om din egen och dina grannars* mängd av brännbart avfall. Sopbilarna väger varje avfallskärl vid tömning. I detta utskick informerar vi om det brännbara avfallets vikt. Läs mer om innehållet i detta utskick på vivab.info/utskickhushallsavfall

Dina egna och dina grannars avfallsvikter under perioden 2019-03-13 till 2019-04-09

Under den aktuella perioden har ditt kärl för brännbart avfall hämtats tre gånger.

Avfallsvikter (kg/person) under den aktuella perioden

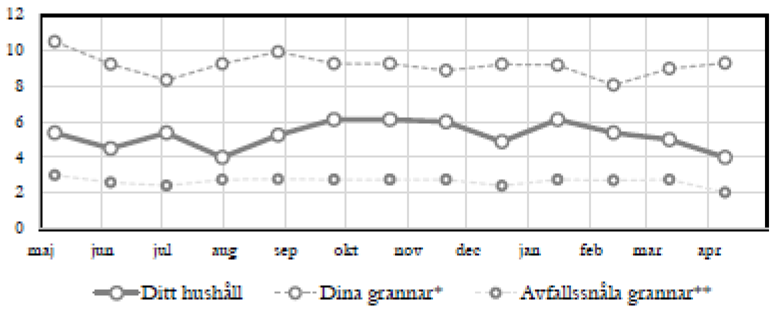


* Genomsnittet baseras på ungefär 100 hushåll i din närhet.
** Den "avfallssnåla" grannen motsvarar genomsnittet av de 20% med minst avfall baserat på ungefär 100 hushåll i din närhet.

Omdöme om ditt hushålls avfallsvikter

UTMÄRKT ☺☺	BRA ☺	KAN FÖRBÄTTRAS

Jämförelse över de senaste 12 månaderna (kg/person)



På vivab.info/utskickhushallsavfall kan du läsa mer om utskicket och om hur personuppgifter hanteras. Här kan du även avregistrera dig från fler utskick inom projektet. Uppge tjänstenumret [redacted]

Figure 1. "Static" feedback in Varberg.

(i) the receiving household's summed residual/unsorted waste weights per person
(ii) average summed per-person weights in a reference group of roughly 100 households in the same treatment arm
(iii) average per-person weights in the subset of "waste efficient" neighbors, i.e. those reference households that belong to the "best" (i.e. lowest) 20 percentiles of the weight distribution

Following what is by now standard practice, we add an injunctive component to the norm with the aim of counteracting the 'boomerang effect', i.e. that efficient households reduce their efforts at the same time that inefficient households increase them (Schultz et at, 2007). Below the bar chart, a summary box with three possible outcomes is displayed. First, if the household weight is above the reference-group average, "Room for improvement" (right) is displayed, with the other two outcomes greyed out. If the household value is below the reference average and the efficient outcomes, "Good" (middle) is displayed instead, along with one smiling emoticon. Finally, if the household value is below the efficient average, "Great" (left) is displayed along with two smileys.

The lower graph shows the evolution of the own-household weight, reference average and efficient average over the past twelve months. Like the upper chart, this time series is updated with each additional feedback letter. Finally, at the bottom of the page is a link to a web page with more information, including some "frequently asked questions". Recipients are also informed that they may opt out from received letters in the future on this web page.

Overall, there are few major differences in the static feedback provided to households in Varberg and Partille. The most substantial difference is that only households in Varberg receive a text evaluation of the bar chart; due to concerns about public acceptance, this feature was removed from the Partille feedback. In Partille, the valenced feedback is reduced to the use of emoticons at the right end of the bar chart, aligned with the upper (own-household) bar. The number of smileys is the same as in Varberg, however, with e.g. one smiley displayed when the household weight lies between the reference average and the efficient average.

Figure 2 shows an example of feedback in the dynamic condition, received only by households in Varberg. Here the time-series graph featured in the static feedback letter is replaced with a centrally placed text box, which reports how waste weights have changed over the immediately preceding four-week period. Thus, the dynamic feedback complements the period-by-period 'snapshot' of how weights are distributed within the reference group with an enhanced focus on the changes that are occurring (Sparkman & Walton, 2017).

The text in Figure 2 reads "During the period in question, your waste weight has decreased by 0.3 kg/person compared with the preceding four-week period. During the same period, 45 percent of your neighbors have reduced their waste by more than your household." Households that increase their waste weight from one month to the next receive similar feedback, but the second sentence now reports the proportion that have reduced their waste at all. We do this to ensure that households are always provided with a relevant benchmark for comparison.

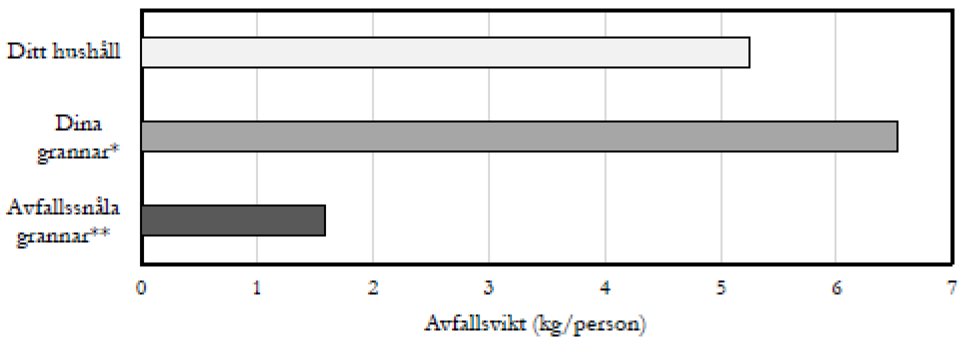
VIVAB strävar efter att minska det brännbara avfallet och vill med detta utskick informera dig om din egen och dina grannars* mängd av brännbart avfall. Sopbilarna väger varje avfallskärl vid tömning. I detta utskick informerar vi om det brännbara avfallets vikt. Läs mer om innehållet i detta utskick på vivab.info/utskickhushallsavfall

Dina egna och dina grannars avfallsvikter under perioden 2019-03-13 till 2019-04-09

Under den aktuella perioden har ditt kärl för brännbart avfall hämtats tre gånger.

> Under den aktuella perioden har din avfallsvikt minskat med 0,3 kg/person jämfört med föregående fyraveckorsperiod.
>
> 45 procent av dina grannar* har under samma period minskat sin avfallsvikt med mer än ditt hushåll.

Avfallsvikter (kg/person) under den aktuella perioden



* Genomsnittet baseras på ungefär 100 hushåll i din närhet.
** Den "avfallssnåla" grannen motsvarar genomsnittet av de 20% med minst avfall baserat på ungefär 100 hushåll i din närhet.

Omdöme om ditt hushålls avfallsvikter

UTMÄRKT ☺☺	BRA ☺	KAN FÖRBÄTTRAS

På vivab.info/utskickhushallsavfall kan du läsa mer om utskicket och om hur personuppgifter hanteras. Här kan du även avregistrera dig från fler utskick inom projektet. Uppge tjänstenumret [redacted]

Figure 2. "Dynamic" feedback in Varberg.

Each feedback letter also includes text on the back, with general information regarding the recycling opportunities in the municipality and some specific tips on how to reduce waste (e.g. by planning food purchases, putting a “no ads” sticker on the mailbox). This page did not vary across time, although there was some variation across the two municipalities. An example back page (for Varberg) is given in Appendix A.

3. Data sources

We will use the following sources of data:

1) A list of all single-family homes served by each municipal waste department, for performing treatment (cluster) randomization, see section 4.

2) Household and collection specific waste weights, for use in (i) construction of norm-based feedback (see section 5) and program evaluation (see sections 6-8). The raw data contains one line per bin-specific collection event; we will sum associated weights across longer time periods and also perform a number of trimming measures, see sections 5 and 6. This data set includes a variable which allows us to identify the length of the waste collection cycle chosen by that household, e.g. once every two weeks.

3) Swedish Tax Authority data on the number of registered residents and each resident’s personal identification number (from which age may be deduced) at all addresses in Partille and Varberg. This data set will be used both when constructing norm-based feedback (section 5) and in data analysis (sections 6-8).

4) A list of households that have opted out of the study, to be collected continuously during the study, see section 5. For reasons of data protection, this list will be handled by a third party and will not be accessed by the authors of the study. Households that have opted out will no longer receive norm-based feedback during the intervention period (section 5), but will be included in subsequent data analysis (sections 6-8). However, their identity at that stage will be unknown to the researchers; thus, no subgroup analysis (e.g. with respect to heterogeneous treatment effects) will be performed on them.

5) Household responses in an online survey, to be sent out during Fall 2019, mainly concerning participant acceptance and effect mechanisms; see Appendix C. Households that have opted out of the study (see above) do not receive an invitation to the survey.

6) Data on the post-intervention waste composition in Partille, to be collected on a total of three days within a single week, during Fall 2019. This data set is expected to include proportions (in terms of weight) for various waste fractions, reported separately for both treatments (i.e. pooled) and control. For this analysis, waste will be collected from (up to) 423 treated households and 238 control households. These are all located in the same part of the municipality and include (at least) a significant share of nine of

the 55 stratification blocks – see section 4. The households involved are selected on the basis of collection lists provided by the municipality; we collect their waste on the same day (i.e. just before) their waste was supposed to be collected under normal circumstances.

7) GPS coordinates for each household and recycling station (for leaving packaging and newspaper waste) in Partille and Varberg.

4. Randomization methodology

The experimental data have a multilevel structure. Prior to the start of the study (during 2018), we received lists of all single-family households served by the municipal waste department in Partille and Varberg, respectively. In Varberg, the list was not a separate document, but was contained in a GIS database covering all relevant households. In either case, starting from the list of households, we first excluded a number of households on the basis of maps and satellite images. In Partille, the excluded group consisted mostly of rural households, while in Varberg, it consisted mainly of households that were clearly of a different type from its immediate neighbors (e.g. farm buildings surrounded by residential areas). In both municipalities, we also excluded addresses that were clearly not households (small businesses, etc.).

The remaining households were manually sorted into contiguous blocks of roughly 100 households each. Care was taken to ensure each block consisted of similar housing types and contained roughly an equal number of households; we attempted to strike a reasonable (though discretionary) balance between these principles whenever they conflicted. Each block was then divided as equally as possible into three contiguous, numbered clusters of about 30 households each (cluster 1, 2, and 3, typically from northwest to southeast). The clusters are the unit of randomization; treatment status is perfectly correlated within cluster. To minimize the risk of contamination between treatment and control, which we hypothesize operates between direct neighbors, we strove to place cluster borders (including across blocks) to minimize the number of direct across-border connections between households. As a result, households on the same street tended to be assigned to the same cluster, and cluster borders tend to run through back yards and green areas, as seen in Figure 1.

The blocks represent a form of stratification by neighborhood, with all three treatment arms represented within each block. As there were three numbered clusters per block, this created six possible permutations of treatment arms across the clusters within each block. We use a random number generator to determine which of these six combinations apply within each block.

In Partille, the above methodology produced 55 * 3 clusters with 34.88 households on average (SD = 3.55 households), a range of 23-43 households per cluster, and a total of 5,756 households. In Varberg, there were 172 * 3 clusters with an average of 30.47 households (SD = 4.63 households), ranging between 17 and 45 households per cluster, and a total of 15,723 households.

Figure 1. Example of design structure. Solid lines are block boundaries, while dashed lines are within-block cluster boundaries. Block number 1 is shown in the upper left, and block 2 in the lower left; see the number in the lower right corner of each block.

5. Study implementation

Households in a “monthly” condition receive a total of nine feedback letters, while households in the “quarterly” condition receive three letters. The four-week (“monthly”) comparison periods coincide in Partille and Varberg, and always run from a Wednesday to a Tuesday. The twelve-week (“quarterly”) comparison period in Partille always corresponds to exactly three four-week periods, thus also running from Wednesday to Tuesday. For example, the first set of both monthly and quarterly letters is compiled on 13 March 2019; thus covering either the four-week period from 13 February to 12 March, or the twelve-week period from 19 December 2018 to 12 March 2019. Subsequent periods have a similar structure.

The timeline for each batch of feedback letters runs as follows. On the Wednesday immediately following the end of the relevant comparison period, the monthly letters (and quarterly, when applicable) are compiled. First a scrubbed data file covering the entire period from March 2018 (i.e. one year before the start of the intervention) up to the current date is constructed using municipal waste data and tax authority information (items 2 and 3 in section 3). This data set has a panel structure, containing: (i) household addresses, (ii) a unique household identifier, (iii) household treatment arm status, (iv) number of waste collection events (successful or otherwise) in each comparison period, (v) feedback information (e.g. own and neighbor per-person waste weights) by comparison period. Whenever an address has no household members in the tax authority data set, we divide total weights by the municipality average, which is 3.0 people/household in Partille and 2.7 in Varberg.

Although feedback information on the behavior of other households is always presented in the letters, no own-household feedback (e.g. own weights) is included in the file (and thus in the letters) if either of two conditions apply. First, if no successful collection event has occurred during the latest comparison period, e.g. due to households never leaving their waste bin out for collection. Second, if at least one collection event is considered unreliable, e.g. due to problems with weighing the bin during collection. We are able to identify such events because the municipal data includes “anomaly reports” associated with some collection events, e.g. when a bin is not placed curbside and thus cannot be collected. Appendix B lists how various anomaly codes are handled in Varberg. In Partille, the list of possible anomalies is less standardized, making a clean summary infeasible. However, our overall coding is highly similar, with some collection events flagged as unsuccessful, some as yielding unreliable weights, and some as being relatively unproblematic such that the research team will ignore the anomaly report.

Furthermore, some households are dropped from the comparison sample during compilation due to noncomparability concerns, e.g. when a property is currently unoccupied. The exclusion criteria are:

- Varberg: the number of two-week periods since March 2018 with no unreliable collection event and at least one nonzero-weight successful collection event (for residual/unsorted waste) is strictly less than three
- Partille: *for either residual and food waste*, no more than 20% of all periods since March 2018 have no unreliable collection event and at least one nonzero-weight successful collection event; or there are collection events of any form in three or fewer periods.

Because of the lack of a 20% criterion, fewer households are typically excluded in Varberg than in Partille. The Partille condition is used because some households consistently produce zero weights for residual waste but strictly positive weights for food waste, and it would then seem that the zero residual weights are “legitimate”, i.e. arising from diligent sorting efforts.

The data file is sent to a third party assistant who matches addresses with recipient names and removes recipients that have opted out. The list of opt-out households is updated immediately prior to this step. The third party then uses the scrubbed waste data to construct a full set of feedback letters. This is printed and mailed to households on the Monday of the following week, implying households receive them on the Tuesday or Wednesday of that week.

6. Data preparation

In our main analysis, we treat the two studies as separate. We first describe how the sample of households is selected. Then, we describe what operations we perform to construct our final data set and the variables used in the analysis. In section 8, we list hypotheses and the statistical analyses designed to test them.

6.1. Sample selection

In each municipality, the sample will be selected as follows.

1) Receive data from the municipal waste department on waste weights for all single-family homes.
2) Retain all households that:
    - Received at least one feedback letter OR
    - Are flagged as part of a control cluster (and thus did not receive any letter), see section 4

This excludes households that were flagged as part of a treatment group but never received a letter, and excludes all single-family homes that were never flagged as part of any of our three treatment arms. We expect a final sample of about 5,500 households in Partille, and about 15,000 households in Varberg.

6.2. Variable construction

The raw waste data sets from Partille and Varberg include three categories of waste bin: food, household (residual), and unsorted waste, where a household typically either has one food and one residual-waste bin, or a single unsorted-waste bin. We recode weight variables associated with residual and unsorted waste as a single residual-waste variable. For each address in the data set, we then perform all the operations described below separately for residual and food waste.

The data sets contain one line per bin-specific collection event, typically associated with a non-missing weight measured in kilograms. However, most households have biweekly collection cycles, with collection for different households roughly evenly spaced across the two weeks. Therefore, in our main analysis, we sum the weights of all events occurring within each two-week interval, as described below. The two-week intervals included in our main analysis are listed in Table 2. Note that these two-week intervals always run from Monday to Sunday at the end of the following week, and therefore do not coincide with the four-week and twelve-week intervals used for feedback purposes, which always run from a Wednesday to a Tuesday.[1]

Recall that events may have associated “anomaly reports” if e.g. a bin was not placed curbside and thus could not be collected. For some such anomaly reports, weights will be recoded as missing values. The coding is identical to that used when compiling feedback letters; see

[1] For example, the first letter sent to households in the monthly condition compared weights throughout the period 13 February-12 March, overlapping completely or partially with periods -2 to 0 in Table 2.

Appendix B for details on how each report is coded in Varberg. In Partille, a similar but much more case-specific coding is performed.

Period	Starting date	Period	Starting date
-25	19 March 2018	-3	21 January 2019
-24	2 April 2018	-2	4 February 2019
-23	16 April 2018	-1	18 February 2019
-22	30 April 2018	0	4 March 2019
-21	14 May 2018	1	18 March 2019 (1st letter received)
-20	28 May 2018	2	1 April 2019
-19	11 June 2018	3	15 April 2019 (2nd letter received)
-18	25 June 2018	4	29 April 2019
-17	9 July 2018	5	13 May 2019 (3rd letter received)
-16	23 July 2018	6	27 May 2019
-15	6 August 2018	7	10 June 2019 (4th letter received)
-14	20 August 2018	8	24 June 2019
-13	3 September 2018	9	8 July 2019 (5th letter received)
-12	17 September 2018	10	22 July 2019
-11	1 October 2018	11	5 August 2019 (6th letter received)
-10	15 October 2018	12	19 August 2019
-9	29 October 2018	13	2 September 2019 (7th letter received)
-8	12 November 2018	14	16 September 2019
-7	26 November 2018	15	30 September 2019 (8th letter received)
-6	10 December 2018	16	14 October 2019
-5	24 December 2018	17	28 October 2019 (9th letter received)
-4	7 January 2019	18	11 November 2019

Table 2. Experimental periods.

Then, in a first stage, for each household we sum all lines across each individual day. In this step, missing weights that occur on the same day as a non-missing weight are dropped from the data set, i.e. summed as zero weights. Lines where (i) the reported weight is negative, or (ii) the event is not associated with any particular waste-bin type are likewise dropped. In the second step, we sum the remaining weights across each two-week interval. Here, any remaining missing values are summed as missing, implying that the two-week period sum will also be missing.[2]

The summed weights are then divided by the number of household members as given by the Swedish Tax Authority. As in section 5, for addresses where the tax authority data does not report any household members, we use the municipality average (3.0 people/household in Partille, 2.7 in Varberg). Furthermore, observations fulfilling the following criteria will be considered outliers and will be dropped from the data at this step:

- All observations of households with a mean residual- or food-waste weight above 15 kg/person

[2] Missing data will not be imputed.

- Each observation where residual- or food-waste weight is above 50 kg/person

Additionally, all observations on households that have more than 90% missing or zero observations (across all two-week periods in Table 2) for both residual and food waste are dropped. We are left with our two household-level outcome variables, for residual and food waste, respectively. We also calculate, for each household, the average residual-waste and food-waste weight across all baseline periods from -25 to 0.

Unless otherwise noted, regressions use treatment variable T_{jt}^{l}, which is always equal to zero for control clusters. For households in letter-receiving treatment arm l (monthly or quarterly letters), $T_{jt}^{l} = 1$ in all periods subsequent to the period of the first letter received; the two-week period ($t = 1$) when the first letter was received is not included in the main analysis. In all other periods, $T_{jt}^{l} = 0$ for these households.

Finally, beyond the baseline averages described above, the set of additional covariates considered in this project are:

(i) Whether the household's waste collection cycle is two weeks or not
(ii) Household size
(iii) Presence of at least one child below five years of age in the household
(iv) Age of head of household. The head of household is interpreted as the oldest member of the household.
(v) Gender of head of household
(vi) Distance D_i, in meters, to the nearest recycling station. The distance variable will be constructed by (i) using household coordinates (see section 3) to calculate, for each household, the distance to every recycling station in that municipality, and (ii) choosing the smallest such distance as D_i.

7. Power calculation

Our power analysis is based on Ek (2019) and starts from the data generating process

$$y_{ijkt} = \lambda_t + \beta T_{jkt} + u_k + u_{kt} + v_{jk} + v_{jkt} + \varepsilon_{ijk} + \varepsilon_{ijkt}$$

which takes blocking into account through the block-specific random terms u_k and u_{kt}. All six error terms are assumed to be i.i.d. with mean zero. From left to right, the (block-, cluster- and unit-related) errors have variance $\sigma_b^2, \sigma_{b\tau}^2, \sigma_c^2, \sigma_{c\tau}^2, \sigma_p^2$, and $\sigma_{p\tau}^2$, respectively. Ek (2019) considers the ANCOVA estimator $\hat{\beta}^{ANCOVA} = \theta\hat{\beta}^{DD} + (1-\theta)\hat{\beta}^{POST}$ used to estimate the average effect of a single treatment versus control. Supposing the weight θ given to the difference-in-differences estimator (as opposed to the ex-post means estimator) is optimal, i.e. minimizes the ANCOVA variance, it can be shown that this variance is

$$Var(\hat{\beta}^{ANCOVA}) = \frac{\hat{\sigma}^2}{rP(1-P)J}\left(1 + (r-1)\hat{\rho} - \frac{mr\hat{\rho}^2}{1 + (m-1)\hat{\rho}}\right) \qquad (1)$$

which is the variance formula we use for power calculation. Here $\hat{\sigma}^2 = \sigma_c^2 + \sigma_{c\tau}^2 + \sigma_p^2/n + \sigma_{p\tau}^2/n$ and $\hat{\rho} = \left(\sigma_c^2 + \sigma_p^2/n\right)/\hat{\sigma}^2$, and n is the average cluster size, as reported in section 4. m is the number of baseline periods (26 in our experiment) and r is the number of endline periods (17 in our main analysis). Finally, P is the share of treated clusters and J is the total number of clusters.

We use the **mixed** command in STATA to estimate all six error terms (Ek, 2019) from data sets covering October 2017-October 2018 in Partille and Varberg, respectively. In these data sets, y_{ijkt} is given by per-person residual waste weights as constructed in section 6.2. Estimating the errors allows us to calculate the ANCOVA variance (1), and corresponding 80% minimum detectable effects (MDE) separately for each municipality. The results are given in Table 3.

Parameter	Partille	Varberg
σ_b^2	0.097	0.127
$\sigma_{b\tau}^2$	0.175	0.122
σ_c^2	0.046	0.000[a]
$\sigma_{c\tau}^2$	0.048	0.042
σ_p^2	10.192	9.050
$\sigma_{p\tau}^2$	6.198	7.262
$\hat{\sigma}^2$	0.573	0.577
$\hat{\rho}$	0.600	0.515
MDE (kgs/person)	0.079	0.050
MDE/mean	2.378%	1.307%

Table 3. Minimum detectable effects corresponding to 80% power in Varberg and Partille.
[a] Note: the ***mixed*** *estimate for this variance component is -0.002, which we round to zero since variances must be positive.*

In both municipalities, nearly all of the variation occurs at the household level; note that the block variance components do not appear in the ANCOVA variance formula above, so the purpose of blocking is simply to 'soak up' variation that would otherwise be seen to operate at the cluster level. Not surprisingly, precision is expected to be higher in the larger Varberg study: the MDE divided by the data average is 1.307% in Varberg and 2.378% in Partille.

In equation (1), suppose we replace the number of clusters J by the number of units N, replace $\hat{\sigma}^2$ by $\sigma^2 = \sigma_c^2 + \sigma_{c\tau}^2 + \sigma_p^2 + \sigma_{p\tau}^2$, and $\hat{\rho}$ by $\rho = \left(\sigma_c^2 + \sigma_p^2\right)/\sigma^2$. This yields the unit-level ANCOVA variance formula given in McKenzie (2012). In fact, under the benchmark assumption of equal-sized clusters, equation (1) is the optimal power formula which would apply (absent, or controlling for, blocking) if the data were first collapsed at the cluster level, replacing unit-level outcomes with cross-sectional cluster averages of those unit outcomes. In this case, $\hat{\sigma}^2$ is the variation in cluster averages, and $\hat{\rho}$ (which is around 50% in both municipalities) is the autocorrelation of cluster averages. Essentially, the McKenzie (2012) formula is applicable under cluster randomization as well, provided cluster averages replace unit outcomes.

Perhaps surprisingly, optimal variance (1) is typically *not* equal to the asymptotic ($J \to \infty$) estimator variance within a standard unit-level ANCOVA regression, i.e. a regression that controls for unit baseline averages, whenever randomization occurs in clusters (Ek, 2019). This is because the regression weight θ generally converges asymptotically to a suboptimal value in unit-level ANCOVA under cluster randomization. With equal-sized clusters, and in line with the previous paragraph, an alternative approach is to collapse the data into cluster averages, on which a cluster-level ANCOVA regression (controlling for baseline averages of such cluster averages) may then be run. This will produce the right asymptotic weight and hence, asymptotically, optimal variance (1).

Nevertheless, when cluster sizes vary and sample size is finite, it is not clear that cluster-level ANCOVA therefore necessarily outperforms unit-level ANCOVA. We have performed simulations suggesting that, at the parameter values that apply to Partille and Varberg,[3] both approaches yield precision acceptably close to that implied by equation (1): rejection rates under nominal 80% power are in both cases at most a few percentage points smaller. Furthermore, any difference in precision between unit and cluster-collapsed ANCOVA is very slight, typically on the order of half a percentage point. We will therefore stick to unit-level ANCOVA in our main analysis (section 8.2), with cluster-level ANCOVA explored as a robustness check (section 8.6).

8. Hypotheses and regression specifications

8.1 Preliminary analyses

Tests of balance: separately for each municipality, we will regress baseline averages and each of the additional covariates listed in section 6.2 on a constant (representing the control-group mean) and a binary indicator for each of the remaining two treatment arms. In addition, we will present difference-in-difference-style “parallel trends” figures showing averages of our main outcome variable, per-person residual-waste weights, for each treatment arm for all periods from -25 to 18.

8.2 Main analysis

All regressions cluster standard errors at the cluster level. Our main regression uses residual-waste weights per person as outcome variable. For each municipality separately, we run the ANCOVA regression

$$y_{ijkt} = \lambda_{kt} + \beta_1 T_{jt}^1 + \beta_2 T_{jt}^2 + \theta \bar{y}_i^{PRE} + \boldsymbol{\gamma} \mathbf{X}_{it} + \epsilon_{ijt} \quad (2)$$

where λ_{kt} are block-by-two week period fixed effects, and $\bar{y}_i^{PRE}$ is the baseline (period -25 to period 0) average of residual-waste weights for household i. Finally, $\mathbf{X}_{it}$ is the full set of

[3] We calculate the coefficient of variation in cluster sizes as the (sample) standard deviation of cluster sizes divided by the sample average, i.e. as $CV = \sqrt{J \sum_j (n_j - n)^2} / \sum_j n_j$, with n_j the number of households in cluster j. In Partille, $CV = 0.1036$; in Varberg, $CV = 0.1482$.

additional controls listed in section 6.2. This regression is run only on data from post-intervention periods, i.e. from period 2 and onwards. Furthermore, it is run both with and without $\mathbf{X}_{it}$. We also run a variant where we do not control for block status, i.e. with time shocks λ_t replacing λ_{kt} in equation 2. In all cases, the null hypotheses of interest are:

Null hypotheses: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_1 = \beta_2$

These are tested separately, with a Bonferroni adjustment applied within regression.

8.3 Checking mechanisms

1) Main data set: same regression equation and treatment coding as above (with and without additional covariates), but with outcome: food-waste weights per person.

   Hypothesis: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_1 = \beta_2$

   These are tested separately, with a Bonferroni adjustment applied within regression.

2) Waste composition data. The proportions of food, combustible, packaging, and paper waste in each sample (treatment vs. control) will be reported. A third-party contractor will perform the composition analysis; if confidence intervals are included in their results, they will be reported in our study.

3) Survey data: for (the sum of) treatments and control, we report point estimates of the share of households reporting whether some major change in waste practices has occurred since the start of the study, as captured by question 3 (see appendix C). We check whether these proportions are significantly different by a chi-square test of differences in two proportions (with df = 1). Next, we report point estimates separately for (the sum of) treatments and control for the various changes made, i.e. question 3.1. Test whether the distribution of changes made differs across pooled treatment/control by performing a chi-square test of homogeneity (since two groups are being compared, this implies df = number of categories – 1). If expected frequency $E_{r,c} < 5$ for any category/treatment arm combination (r, c), that category will be combined with another to ensure the chi-square approximation is adequate.

8.4 Checking for interference/contamination/spillovers

1) Main data set, Partille only: run the following household-level ANCOVA regression

$$y_{ijt} = \lambda_{kt} + \beta_1\left(T_{jt}^1 + T_{jt}^2\right) + \beta_2 S_{ij}^{1,2} + \theta \bar{y}_{ij}^{PRE} + \epsilon_{ijt}$$

   where $S_{ij}^{1,2}$ is a binary variable which we use to flag households that are in the control group, but are adjacent to other households which are in either of the treatment groups; note that the two treatments are pooled in this regression. Adjacency is determined through visual inspection of maps and satellite images and requires a common border between the land plots of the two households or that the only barrier between the properties is a street.

Hypothesis: $\beta_2 = 0$

2) Main data set, Varberg only: run the following ANCOVA regression

$$y_{ijt} = \lambda_{kt} + \beta_1(T_{jt}^1 + T_{jt}^2) + \beta_2 R_j(T_{jt}^1 + T_{jt}^2) + \theta \bar{y}_{ij}^{PRE} + \epsilon_{ijt}$$

where the binary variable R_j is equal to one if the cluster is flagged as a rural cluster, and zero if it is flagged as an urban cluster. Urban/rural status is based on the Statistics Sweden definition of an 'urban center' as a connected residential area housing at least 200 inhabitants. We visually inspect a map of the municipality to find the blocks corresponding most closely to the set of urban centres with at least 300 inhabitants as of the 31 December, 2018, thus making urban status slightly more restrictive than the original definition. Note that, again, both treatments are pooled in the above regression.

Hypothesis: $\beta_2 = 0$

3) Survey data: report summary statistics (share of answers for each alternative) on question 4 (including nested subquestions), pooled across all responding households.

8.5 Heterogeneous treatment effects

1) Main data set: for each municipality separately, use the machine learning approach of Athey & Imbens (2016) at the household level to construct a partition of the covariate space across which treatment effects differ. We use the full set of covariates listed in section 6.2; we may also include average household weights across all baseline periods (-25 to 0), depending on whether or not this appears mostly to capture a "regression-to-the-mean" effect.

2) Regression discontinuity with respect to receiving smileys (see Allcott, 2011; Costa & Kahn 2013, Online Appendix). All steps of this analysis will be performed separately for each municipality. Preliminary regression: include only forcing variable, i.e. run the following

$$y_{ijt} = \lambda_{kt} + \beta_1\left(T_{jt}^1 + T_{jt}^2\right) + \beta_2 D_{ijt}\left(T_{jt}^1 + T_{jt}^2\right) + \theta \bar{y}_{ij}^{PRE} + \epsilon_{ijt}$$

where D_{ijt} is the distance (positive or negative) to the household's reference group average, i.e. to the Room for improvement/Good threshold, given in the feedback letter that is the most recent at time t. At this stage, the null hypothesis is $\beta_2 = 0$.

For the regression discontinuity analysis, we will first plot the data average y_{ijt} as a function of D_{ijt} around (i) the reference group average, and (ii) around the efficient-neighbor average, i.e. around the Good/Great threshold. Although we may adapt our approach depending on what the graphs show, we will at least run the following local linear regression:

$$y_{ijt} = I(0 < |D_{ijt}| < d) * \left(\alpha + \beta I(D_{ijt} > 0) + \gamma_1 D_{ijt} I(D_{ijt} < 0) + \gamma_2 D_{ijt} I(D_{ijt} > 0)\right) + \lambda_{kt} + \theta \bar{y}_{ij}^{PRE} + \epsilon_{ijt}$$

using only observations for which $T_{jt}^1 + T_{jt}^2 = 1$. Here d is the 'bandwidth' around the reference-group average (we will step through various values), and $I(\cdot)$ is a binary indicator function. Through parameters γ_1 and γ_2, we allow for different linear trends on either side of the threshold where $D_{ijt} = 0$. Note that households outside the bandwidth are still included in the regression in order to identify λ_{kt} and θ.

Here, the null hypothesis is that $\beta = 0$. We also run an equivalent regression where D_{ijt} is replaced with the distance to the Good/Great threshold.

8.6 Robustness tests:

1) Cluster-collapsed regression. The data are first collapsed at the cluster level, calculating cross-sectional cluster averages $\bar{y}_j^{PRE}$ of residual-waste weights across all households with non-missing weights. Then, use same regression equation and treatment coding as in section 8.2 (without additional covariates), i.e.

$$y_{jkt} = \lambda_{kt} + \beta_1 T_{jt}^1 + \beta_2 T_{jt}^2 + \theta \bar{y}_j^{PRE} + \epsilon_{jt}$$

This ANCOVA regression is then repeated for food waste, aggregated at the cluster level. In both cases, the null hypotheses are:

Hypothesis: $\beta_1 = 0, \beta_2 = 0, \beta_1 = \beta_2$

which are tested separately, with a Bonferroni adjustment applied within regression.

2) Log waste weights. We perform the adjusted logarithmic transformation $\ln(x+1)$ of all per-person weights x and likewise calculate, for each household, the average adjusted-log residual-waste and adjusted-log food-waste weight across all baseline periods from -25 to 0. Then, we re-run the main regression (2), with the same hypotheses and adjustments.

3) Tobit regression. As in item 2) above, except the regression run is a Tobit regression accounting for truncation at $\ln 1 = 0$. Hypotheses and adjustments are unchanged.

4) Include only households with a two-week collection cycle. For each municipality separately, perform main regression (equation 2) above but use only households with a two-week waste collection cycle. Our waste data set includes information on waste cycle length; see section 3.

5) Include period $t = 1$. For each municipality separately, perform main regression (1) with period $t = 1$ retained in the data set. In this period, for households in letter-receiving groups T_{it}^1 and T_{it}^2 are equal to 0.8, the share of weekdays occurring after the first letter is assumed to have been received (19 March 2019).

6) Pool both municipalities and run the following adjusted main regression:

$$y_{jkt} = \lambda_{kt} + \beta_1 T_{jt}^1 + \beta_2 T_{jt}^2 + \beta_3 T_{jt}^3 + \theta \bar{y}_j^{PRE} + \epsilon_{jt}$$

where T_{jt}^1 now represents the monthly static norm feedback in *both* municipalities, while T_{jt}^2 and T_{jt}^3 are dummies for the monthly dynamic norm feedback treatment and the quarterly static feedback treatment, respectively.

Null hypotheses: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$, $\beta_1 = \beta_2$, and $\beta_1 = \beta_3$

These are tested separately with a Bonferroni adjustment.

7) Repeat all regressions in subsections 8.3 and 8.5 with both municipalities pooled. This excludes the analysis of survey and waste composition data in section 8.3, but includes the machine learning approach to heterogeneous treatment effects in section 8.5. When applicable, null hypotheses will be adjusted to account for the presence of three treatments instead of two. In addition, whenever both (municipality-specific) treatments were summed as a single variable, we now sum all three treatments as one variable.

8.7 Cost-benefit analysis:

Finally, we will also perform a partial or full cost-benefit analysis where we compare the environmental benefits of any waste reduction with direct and indirect costs of the intervention. This analysis should be considered exploratory, and we do not specify details at this stage.

References

G. Alberts, Z. Gurguc, P. Koutroumpis, R. Martin, M. Muûls, and T. Napp. Competition and norms: A self-defeating combination? *Energy Policy*, 96, 2016.

H. Allcott. Social norms and energy conservation. *Journal of Public Economics*, 95, 2011.

H. Allcott. Site selection bias in program evaluation. *Quarterly Journal of Economics*, 130(3), 2015.

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

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

M. A. Andor, A. Gerster, J. Peters, and C.M. Schmidt. Social norms and energy conservation beyond the US. Ruhr Economic Papers No. 714, 2018.

S. Athey and G. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences,* 113, 2016.

I. Ayres, S. Raseman, and A. Shih. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, 29(5), 2012.

D.L. Costa and M.E. Kahn. Energy conservation ”nudges” and environmentalist ideology: evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3), 2013.

P. Dolan and R. Metcalfe. Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation. Working paper, 2015.

C. Ek. A formula for power calculation in cluster-randomized experiments with panel data. Working paper, 2019.

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

M.M. Jaime Torres and F. Carlsson. Direct and spillover effects of a social information campaign on residential water-savings. *Journal of Environmental Economics and Management*, 92, 2018.

J. LaRiviere, S. Holladay, D. Novgorodsky, and M. Price. Prices vs. nudges: A large field experiment on energy efficiency fixed cost investments. Working paper, 2014.

D. McKenzie. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 2012.

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

G. Sparkman and G.M. Walton. Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science* (2017), 28(11), 2017.

APPENDIX A. FEEDBACK LETTER, SECOND PAGE

Bidra till en bättre miljö – ge ditt avfall nytt liv

Det mesta av vårt avfall kan bli material i nya produkter eller återvinnas som energi. Vi vill göra det lätt för dig att sortera, återanvända och återvinna så mycket som möjligt.

Varje sorterad förpackning gör skillnad

I stort sett alla gamla förpackningar kan återanvändas som material i nya produkter. Om alla hushåll återvann en plastförpackning till i månaden skulle koldioxidutsläppen minska med 3600 ton. Det motsvarar utsläppen från cirka 1200 bilar varje år.

Återvinningscentraler i Varberg

Källsorterat avfall som lämnas på våra tre bemannade återvinningscentraler tar vi hand om på olika sätt beroende på material. Hårdplast till exempel, återvinns och blir till råvara i nya plastprodukter, medan tryckimpregnerat trä räknas som farligt avfall och förstörs i särskilda förbränningsanläggningar där energi kan utvinnas.

Återvinningsstationer i Varberg

Du kan lämna tidningar och glas-, kartong-, plast- och metallförpackningar på någon av de 25 återvinningsstationer som finns i Varberg. Förpacknings- och tidningsinsamlingen, FTI, ansvarar för behållarna och att de töms regelbundet. På ftiab.se kan du hitta en återvinningsstation som ligger nära dig.

Tips från Avfall Sverige för att minska ditt avfall

FÖREBYGGANDE

- Undvik att behöva slänga mat som blivit gammal genom att planera dina matinköp.
- Slipp oadresserad direktreklam genom att sätta upp en "nej-tack-till-reklam-skylt" på brevlådan.
- Låna eller dela redskap och verktyg som du använder sällan med dina grannar och vänner.

ÅTERANVÄNDNING

- Ta med tygpåsar eller gamla pappers- eller plastpåsar att packa dina varor i när du går till affären.
- Lämna textilier, inredning och byggmaterial på kommunens återvinningscentral. Du kan också köpa och sälja på secondhandsajter på internet.
- Lämna in trasiga saker för reparation istället för att köpa nytt.

MATERIALÅTERVINNING

- Lämna förpackningar och tidningar på din närmaste återvinningsstation.
- Välj ett sophämtningsabonnemang med utsortering av matavfall. Eller kompostera ditt organiska avfall själv.
- Fler återvinningstips finns på ftiab.se/sortering.html

För mer information se vivab.info

Figure A.1 Second page of Varberg feedback letter.

APPENDIX B. ANOMALY REPORTS (VARBERG)

	Report code	Coding in data
Bin not curbside	010	Unsuccessful (0)
Blocked, car	020	Unsuccessful (0)
Blocked, snow	030	Unsuccessful (0)
Blocked, other	040	Unsuccessful (0)
Locked door/gate	050	Unsuccessful (0)
Not shoveled	060	Unsuccessful (0)
Not plowed	070	Unsuccessful (0)
Not gritted	080	Unsuccessful (0)
Incorrect bin contents, not collected	090	Unsuccessful (0)
Incorrect bin content, collected	095	Ignore
Overfull	100	Ignore
Heavy bin	105	Ignore
Other	110	Ignore
Broken bin	120	Ignore
Bar code missing	130	Unreliable (missing)
Label missing	135	Ignore
Empty bin	140	0
Sacks collected	150	Unreliable (missing)
Broken wheel	160	Ignore
Food waste bag	165	Ignore
Food waste bags often	166	Ignore
Broken lid	170	Ignore
Cannot find bin	180	Unsuccessful (0)
Bar code broken	190	Unreliable (missing)
Manual collection	195	Unreliable (missing)

Table 2. Anomaly report codes

APPENDIX C. TRANSLATED POST-INTERVENTION SURVEYS

Appendix C.1 presents the full questionnaire sent to treated households in Varberg. The questionnaire is implemented online through the Qualtrics software and an invitation with a link and a QR code sent out as part of the last feedback letter (treated households) or a separate letter (control households). Control households face a similar but shorter survey to that in Appendix C.1, containing only questions 1, 3, 4 (excl. 4.3), and 5. With only very minor variations, the same questionnaires are sent to treated and control households in Partille.

Appendix C.1 Survey to treated households in Varberg

What do you think?

Over the last few months, you have received letters from VIVAB containing information on your amounts of combustible waste. We have aimed to examine whether people are motivated to reduce their waste when they become aware of its weight in comparison with that of their neighbors.

We want to know what you think! We would like you to answer a few questions about the letters and about the project in this questionnaire.

The survey takes 2-3 minutes to complete.

1. Have you been aware of the project during the last few months?

   ☐ Yes

   ☐ No

2. Do you think that the letters have helped you to reduce your combustible waste?

   ☐ To a large extent

   ☐ To some extent

   ☐ Not at all

3. Have you made any changes in how you manage waste since March 2019?

   ☐ Yes
   ☐ No

3.1 If yes, what is the most significant change that you have made? [Single answer required]

I have…

☐ put more containers in my home for sorting waste

☐ improved at sorting packaging and paper

☐ improved at sorting food waste

☐ started to plan my purchases so as to generate less food waste

☐ started to plan my purchases so as to generate less packaging waste

☐ Other: ____

4. Have you discussed the letter with people in other households?

☐ Yes
☐ No

4.1 If yes: who have you mainly discussed them with? [Single answer required]

☐ Neighbors living adjacent to my property

☐ Neighbors not living next to my property but on the same street

☐ Other neighbors in the vicinity (within a radius of 100 metres)

☐ Other people

4.2 If yes: did you discuss different ways of reducing your waste?

☐ Yes

☐ No

4.3 If yes: did you disclose your weights to each other?

☐ Yes

☐ No

5. Would you like to receive more information with waste comparisons in the future?

☐ Yes

☐ No

5.1 If yes: how often would you like to receive the waste comparisons?

☐ Once a month

☐ Once every three months

☐ Once every six months

☐ Once a year

5.2 If yes: how would you like to receive the waste comparisons?

☐ Through a waste services mobile app

☐ As paper mailings

☐ On the invoice for waste services

☐ Through “My pages” on vivab.info

6. In general, what did you think of the letters?

☐ Liked them a lot

☐ Liked them somewhat

☐ Neither liked nor disliked them

☐ Disliked them somewhat

☐ Disliked them a lot

7. Was there some information in particular that you think the letters lacked? [Answer not required]
