Working Paper in Economics No. 804

Norm-based feedback on household waste: Large-scale field experiments in two Swedish municipalities

Claes Ek, Magnus Söderberg

Department of Economics, May 2021

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Norm-based feedback on household waste: Large-scale field experiments in two Swedish municipalities[☆]

Claes Ek[a,1], Magnus Söderberg[b,c]

[a] *Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden*
[b] *Department of Sociology, Environmental and Business Economics, University of Southern Denmark, Niels Bohrs Vej 9-10, DK-6700 Esbjerg, Denmark*
[c] *Ratio Institute, P.O. Box 3203, SE-103 64 Stockholm, Sweden*

Abstract

We conduct separate randomized controlled trials of norm-based feedback nudges on household waste in two municipalities in western Sweden. Our main treatment presents recipients with accurate, household-specific feedback highly similar to the standard Home Energy Report design, but with residual (unsorted) waste as the object of comparison. We also test a novel 'dynamic' norm design informed by psychological research. Post-experimental reductions are on the order of 7-12% in both municipalities, substantially larger than in most previous studies. We estimate that the reduction corresponds to a 30-60% increase in unit-based waste fees. Effect differences between our main treatment and the dynamic-norm treatment are not significant. We find that feedback nudges are highly cost-effective compared to alternative means for reducing household residual waste. However, net social benefits depend on whether existing waste fees internalize the marginal social cost of residual waste. Our results have implications for the usefulness of feedback interventions as well as for unit-based pricing of waste, on which our feedback materials rely.

Keywords: Field experiments, household waste, norm-based feedback, unit-based pricing, pay-as-you-throw

JEL classification: D13, I21, Q53

[☆] This version: 17 May 2021. This project was funded by the Swedish research council Formas, grant no. 2017-00225. Ruben Dieleman provided research assistance.

Email address: claes.ek@economics.gu.se (Claes Ek)

[1] Corresponding author. Tel.: +46 73 723 73 86.

1 Introduction

Large-scale interventions promoting household resource conservation through norm-based feedback have become a mainstay of applied behavioral economics over the past decade.[2] The most well-known example of such a norm-based intervention is the 'Home Energy Report' (HER) developed by Opower and mailed to households across the United States (see e.g., Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013). HERs present household-specific feedback on energy use compared to a set of similar neighbors, and includes both injunctive ('ought') and descriptive ('is') norm components.

Effects from norm feedback interventions appear to systematically differ in size across contexts and domains. Feedback on water use appears to drive reductions of about 5% in receiving households (Ferraro and Price, 2013; Bernedo et al., 2014; Jaime Torres and Carlsson, 2018). Effect sizes for electricity use are generally smaller: in an analysis of over 100 different large-scale HER experiments in the US, Allcott (2015) finds average reductions of 1.31%, with a standard deviation of 0.45 percentage points. Given the documented range of effects, there has recently been some debate on the value and cost-effectiveness of feedback interventions in different settings.[3] Our paper adds to that discussion by applying the HER paradigm in a novel domain: household waste. We are aware of no other study that tests the large-scale impact of household-level norm feedback on waste.

It seems plausible that the range of observed effect sizes reflects differences in how much cost and effort is required to reduce usage. If so, there is good reason to expect larger effects from feedback on waste than on either electricity or water use. According the 2015 Residential Energy Consumption Survey of the US Energy Information Administration,[4] the bulk of home electricity use is for air conditioning, refrigerators, and space and water heating, with only about 10% due to lighting. Making deep cuts in household electricity use thus involves either high-effort behavior change such as turning down the heat in winter, or

[2] The rapidly growing literature, starting from Schultz et al. (2007), now includes a large number of studies evaluating specific designs (Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013; Ferraro and Price, 2013; Dolan and Metcalfe, 2015; Jaime Torres and Carlsson, 2018; Holladay et al., 2019; Brülisauer et al., 2020), as well as long-run effects (Ferraro et al., 2011; Allcott and Rogers, 2014; Bernedo et al., 2014), welfare implications (Allcott and Kessler, 2019), and psychological mechanisms (Alberts et al., 2016; Byrne et al., 2018).

[3] In particular, Andor et al. (2020) has recently argued that the relative success of HER in the US may be limited to that setting. Allcott (2011) estimates that, for effect sizes of about 1-3%, the Opower HERs are cost-effective per unit of carbon emissions compared to other energy conservation policies. By contrast, Andor et al. (2020) replicate the HER design in Germany and observe a substantially smaller treatment effect of 0.7%. Since baseline per-capita electricity consumption is higher in the US than in other OECD countries, the authors conclude that norm-based feedback is unlikely to be cost-effective outside of the United States.

[4] `https://www.eia.gov/consumption/residential/index.php`.

costly physical capital investments. Allcott and Rogers (2014) observe that few households respond to feedback by buying energy efficient appliances, suggesting limited scope for low-cost behavior change. A similar, but less stark pattern seems likely to hold for water use (Bernedo et al., 2014). By contrast, reducing unsorted waste normally requires little in the way of physical capital investment beyond buying a set of in-home recycling bins. Unlike for electricity or water, this physical capital is largely a complement rather than a substitute with effort (Vollaard and van Soest, 2020).[5] Even so, the marginal effort of increasingly diligent recycling seems less steep compared to reductions in other domains, and household waste may thus represent an upper bound on what HER-style feedback can achieve.

Our results are consistent with these points. We run separate large-scale experiments in two municipalities in western Sweden. Randomizing treatment across nearly all single-family homes in both localities, we send treated households repeated and accurate feedback, presented in a HER-style format, on the amount of residual waste (in kg/person) they generate compared to neighbors.[6] We focus on residual waste rather than recycling rates to allow for waste prevention, for example by buying less packaging-intensive consumer products. In both experiments, high-precision average treatment effect (ATE) estimates show that the residual-waste weights of treated addresses drop after receiving the first letter.

Notably, reductions are about 7-12%, depending on the exact regression specification: considerably larger than in previous studies of norm-based feedback. In fact, the reduction is about the same magnitude as in a recent study by Vollaard and van Soest (2020) on a Dutch crackdown on incorrect sorting involving both fines and salient bin inspections. In principle, the effect we observe may be driven by any combination of waste prevention, increased recycling, and illicit disposal (dumping). When examining mechanisms, we find no evidence that treatment increases illicit disposal. Thus, given that our main treatment replicates the standard HER format, our results confirm that effects from feedback do indeed vary strongly across domains; moreover, it suggests that norm-based interventions may be deployed as effective non-price instruments to reduce unsorted waste.

Both participating municipalities have pre-existing systems for unit-based pricing (UBP) of waste (weight-based fees). Indeed, our particular design is explicitly tailored to such

[5]Exceptions involving substitutability with effort do exist and tend to involve actions to prevent waste, e.g., by placing a ‘no ads’ sticker on one’s mailbox.

[6]Throughout this paper, the term residual waste is used for the unsorted fraction of household waste, which is typically incinerated in most OECD countries (including Sweden). Similarly, what we term food waste is the biodegradable fraction collected by utilities, which differs from wasted food in that not all food waste is avoidable.

pricing schemes: our partner utilities routinely weigh each bin during collection, and feedback letters are constructed from the resulting weight data. As a result, our study also relates to the ongoing examination of marginal-cost pricing of waste, also known as 'pay-as-you-throw'.[7]

As in the case of energy or water, utilities may be reluctant to raise marginal costs further, e.g. because of acceptance concerns. Additionally, there is some worry that higher prices per unit will lead to increased dumping of waste (Fullerton and Kinnaman, 1996; Heller and Vatn, 2017) or leakage to unregulated areas ('waste tourism'; see e.g., Bucciol et al., 2015), though the weight of the empirical evidence suggests such perverse effects on disposal are small to nonexistent (Allers and Hoeben, 2010; Carattini et al., 2018; Bueno and Valente, 2019; Erhardt, 2019; Valente, 2020). Nevertheless, non-price policies may prove more readily implementable than further increases in the per-unit price facing households.[8] Conversely, the fact that there is unique potential to add norm-based interventions on top of pay-as-you-throw schemes implies that UBP may hold greater promise than has yet been recognized. We estimate that unit-based fees need to increase by 32-60% to produce effects of similar size to our main ATEs. Thus, norm feedback appears highly effective even compared with economic incentives.

Feedback is also very cost-effective compared to other non-price waste policies such as curbside collection of packaging. Impacts on social welfare are more mixed: in one of the two municipalities, we find that existing unit-based fees already account for most social costs of waste disposal, so net benefits from feedback are unsurprisingly negative. By contrast, marginal fees are lower in the other municipality, implying generally positive net social benefits of up to about \$20 per household. Thus, feedback nudges stand out as efficient second-best policy for areas where pre-existing marginal incentives do not (fully) reflect

[7]Early studies of UBP have attempted to identify the causal effect of such schemes on waste generation and recycling by making before-after comparisons (Fullerton and Kinnaman, 1996), exploiting cross-sectional variation (Kinnaman and Fullerton, 2000), or both (Dijkgraaf and Gradus, 2009). More recently, several studies have used a regression difference-in-differences approach with unit (e.g., municipality) fixed effects (Allers and Hoeben, 2010; Usui and Takeuchi, 2014; Bucciol et al., 2015; Dijkgraaf and Gradus, 2017; Carattini et al., 2018). Bueno and Valente (2019), arguably the current 'state of the art' in the UBP literature, uses a synthetic-control strategy that appears to better model unobserved heterogeneity than fixed-effects approaches do. Finally, some studies have also used complementary IV approaches to control for endogenous policy (Kinnaman and Fullerton, 2000; Allers and Hoeben, 2010; Huang et al., 2011).

[8]In line with this point, several authors have noted that recycling efforts appear strongly driven by intrinsic motivation, in addition to material concerns (e.g., Sterner and Bartelings, 1999; Berglund, 2006; Kipperberg, 2007; Ferrara and Missios, 2012; Czajkowski et al., 2017). Interestingly, Viscusi et al. (2011) question the importance of social norms for waste behavior, arguing that private values are crucial; in contrast, our results would seem to confirm that waste behavior is strongly driven by norm-related concerns.

marginal social costs. At the end of this paper, we also discuss potentials for applying waste feedback in areas without *any* unit-based pricing.

A final contribution of our study is that, in one of the two experiments, we test not only the standard HER feedback format but also a novel design that stresses how resource use in a household's comparison group changes over time. Recent evidence in Sparkman and Walton (2017) and Mortensen et al. (2019) suggests that such 'dynamic' or 'trending' norms are more effective at changing behavior. They argue that, compared with snapshot information stressing cross-sectional variation, presenting respondents with ongoing changes provides a stronger signal that behavioral costs and benefits are shifting in the population as a whole. Thus, respondents may be more inclined to respond by updating their view of the proper course of action. We do identify substantial waste reductions also among households receiving dynamic-norm feedback. However, the effect is statistically indistinguishable from the standard HER design: thus, we find no evidence that dynamic-norm feedback interventions are more effective.

The remainder of this paper is organized as follows. In Section 2, we provide some brief institutional background on waste management in Sweden. Section 3 outlines our experimental design, while Section 4 describes our empirical strategy as well as some important features of the data. Section 5 presents main results. Section 6 then moves on to various extensions, analyzing mechanisms (illicit disposal, prevention, recycling), treatment-effect heterogeneity, the effect of injunctive labeling, and long-run effects. Section 7 evaluates the costs, benefits, and cost-effectiveness of norm feedback on household waste. Finally, Section 8 concludes the paper.

2 Waste management in Sweden

Swedish national targets for waste management largely derive from EU objectives, with the 2018 revision of the EU Waste Framework Directive requiring each members state to recycle 50% of household waste by 2020. Additional targets construct a trajectory where recycling targets increase by five percentage points every five years, up to 65% in 2035. The overall Swedish recycling rate stood at 57% in 2018, so the 2020 target is being met, though additional policies are needed to attain later targets. There are also more specific Swedish targets for packaging, paper, and food-waste recycling, not all of which are currently being met (Swedish Environmental Protection Agency, 2020).

Local waste management rests on a dual system. First, collection and treatment of residual and food waste is left to municipalities, typically being run by local utilities. Some localities have opted for curbside collection and/or unit-based pricing to encourage household recycling; the latter is in use in about 10% of Swedish municipalities. Second, packaging and paper are subject to extended producer responsibility regulations. Collection from single-family homes (the focus of our study) occurs mainly through some 5,000 designated 'recycling stations' where households may go to drop off packaging and paper waste. All stations are run by a single producer-owned corporation, FTI.[9]

Our experiments were conducted in Varberg and Partille, two municipalities in southwest Sweden. Both have used weight-based waste fees since the 1990s, and neither collects residual or food waste curbside. In both municipalities, waste fees have a fixed as well as a per-unit component. The marginal-cost component remained constant throughout 2019, the year of our intervention; in Partille, it equalled approximately \$0.20 per kg in USD terms, for both residual and food waste, while it was about \$0.34 in Varberg. Varberg additionally requires households that do not source separate food waste to pay a per-unit surcharge, roughly doubling the per-unit price. In both areas, the variable cost component is displayed separately on all utility bills received by households.

3 Experimental design

We conduct a pair of separate but parallel studies in the Swedish municipalities of Varberg and Partille. In each locality, our study sample includes about 90% of all single-family homes; since household-specific waste weights cannot be identified in apartment buildings, no such addresses are part of either study. This leaves us with about 15,000 households in Varberg and 5,000 in Partille.

In both areas, households are divided roughly equally into three treatment arms, including a control group; however, the two treatments differ across municipalities, as shown in Table 1. All experimental interventions involve letters containing accurate and household-specific norm feedback on residual waste. These letters, stamped with the relevant municipal logo, are sent repeatedly to all treatment-group households. Both interventions took place during March-October 2019, with the first letters received on 19 March in all groups.[10]

Households in either control condition do not receive feedback letters. As for treatments, first, the Varberg study varies the feedback type used. We attempt to go beyond stan-

[9]Citing dissatisfaction with how the recycling stations are managed, some municipalities now offer curbside collection of packaging and paper in addition to residual and food waste. The two municipalities we study do not, however.

[10]There is variation of up to one day around all receiving dates because of limitations in the delivery capacity of the Swedish postal service.

Municipality	Treatment	N
Varberg	1. *Control*: no letters sent	4,971
	2. *Static*: monthly norm-based feedback	4,961
	3. *Dynamic*: monthly norm-based feedback	5,003
Partille	1. *Control*: no letters sent	1,837
	2. *Monthly*: 'static' norm-based feedback	1,838
	3. *Quarterly*: 'static' norm-based feedback	1,844

Notes: Table lists treatment conditions in the two studies. For each treatment, the final column reports how many addresses are included in our main data sets.

Table 1: *Experimental treatments in the two studies*

dard HER-type feedback designs emphasizing cross-sectional comparisons between households ('static') to instead highlight how waste behavior has changed since the last letter was received ('dynamic'). In Partille, all households receive static feedback, and we instead vary feedback frequency, 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 between March and October 2019, while households in the 'quarterly' condition receive three feedback letters.

Figure 1 provides an example, translated from Swedish, of the 'static' feedback presented to households in Varberg. The setup is very similar to previous studies on HERs such as Allcott and Rogers (2014) or Andor et al. (2020). In monthly conditions, each letter refers specifically to the preceding four weeks; in the quarterly condition, reference periods are the past twelve weeks. For each such period, the bar chart in the upper part of the page displays, top to bottom: (i) the receiving household's summed residual-waste weights per person; (ii) average summed per-person weights within a reference group of roughly 100 households belonging to the same treatment arm; and (iii) average per-person weights within the subset of 'waste efficient' neighbors, i.e., households in the bottom 20 percentiles of the reference-period specific weight distribution.[11]

Following standard practice, we add an injunctive component to the bar chart, with the

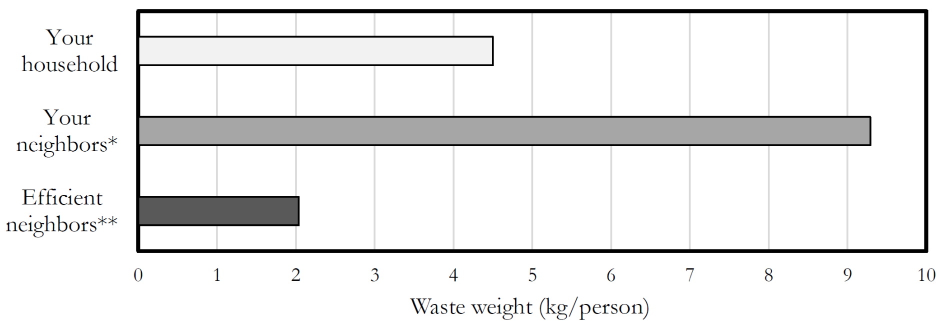
[11] Other studies have made the comparison with the 20th percentile instead; we believe the average is easier to explain. A second difference is that we did not include information on potential monetary savings.

VIVAB works to reduce the amount of unsorted waste. With this letter, we want to provide information on the amount of unsorted waste generated by you and your neighbors*. Our garbage trucks weigh each bin during collection. The information in this letter concerns the weight of the unsorted waste. For more information on the contents of the letter, see vivab.info/utskickhushallsavfall

Your own waste and your neighbors' waste, during the period 2019-03-13 to 2019-04-09

During this period, your bin for unsorted waste has been collected three times.

Waste weights (kg/person) during the period

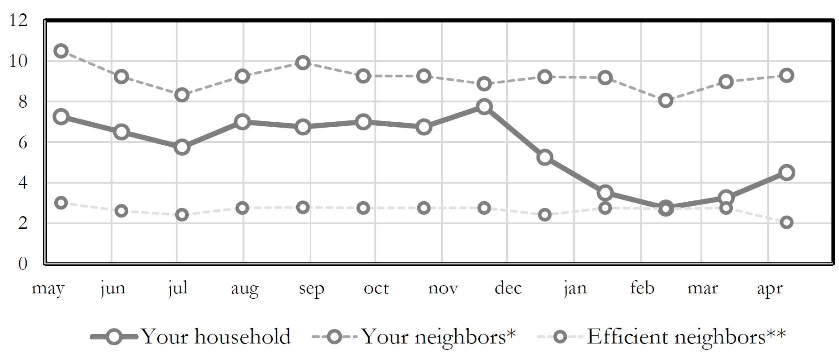


* The average is based on about 100 households in your neighborhood.
** This 'waste efficient' neighbor represents the average among the 20% of households with the least waste, based on about 100 households in your neighborhood.

Evaluation of your waste weight

EXCELLENT ☺☺ GOOD ☺ ROOM FOR IMPROVEMENT

Comparison over the last 12 months (kg/person)



For more information about the letters and how we process your personal data, see vivab.info/utskickhushallsavfall. You can also go there to opt out from receiving additional letters within the project. Please state your service ID: XXXXXX

Figure 1: Translated example of a static-norm feedback letter

aim of counteracting ‘boomerang effects’, i.e., that efficient households reduce their efforts at the same time that inefficient households increase them (Schultz et al., 2007). Below the bar chart, a summary box with three possible outcomes is displayed. First, if a household’s weight is above the reference-group average, the assessment ‘Room for improvement’ is displayed, with the other two outcomes greyed out. For weights below the reference average but above the efficient average, ‘Good’ is displayed instead, along with one smiling emoticon. Finally, if the weight falls below even the efficient average, ‘Great’ is displayed along with two smileys.

The lower graph shows the evolution of own-household weights as well as reference and efficient averages 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 municipal web page with more information, including some ‘frequently asked questions’. A translated version of such an FAQ section may be found in Appendix A.1.[12] Recipients are also informed that the FAQ web page includes a service where they may opt out of receiving letters in the future. Households that do so receive no further feedback letters during the entire intervention period.

By the end of the project, 1,466 households had opted out in Varberg, while 189 households had done so in Partille. These figures amount to 14.6% and 5.1% of treated households, respectively; by comparison, in studies of HERs, opt-out rates tend to be less than 1% on average (Allcott, 2015). Some households are likely to view simply being exposed to the letters as a utility cost (Allcott and Kessler, 2019); while not visible in our data, the high opt-out rates we observe suggest the magnitude of such costs may be domain-specific, i.e., larger than for energy-use feedback. In any case, although households that opted out are also not identifiable as such in our final data set, we do retain them in the sample, and thus their decision not to participate does not bias our results in relation to actual policies with similar opt-out rates.

Overall, few major differences exist between the static feedback received by households in Partille and Varberg. The most substantial difference is that, due to municipality concerns regarding public acceptance, households in Partille do not receive a textual evaluation of the bar chart. 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, for example with one smiley displayed when the household weight lies between the reference average and the efficient average. However, we do not grey out the set of possible

[12] Among other things, the FAQ section stresses that the letters are for information provision only and that high-waste households will not face sanctions.

assessments not given to a household.

Figure 2 shows a translated 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. For households that have reduced their weight, the share of neighbors with an even larger reduction is given. Households that increase their waste weight from one period to another receive similar feedback, but with the sentence on neighbor behavior reporting the proportion that have reduced their waste by any amount. Thus, households are always provided with a relevant benchmark for comparison.

It is worth noting that 'static' letters do include a (dynamic) time series; likewise, even in the 'dynamic' letters, the text-and-emoticon evaluation still refers to the (static) bar chart. Thus, the static/dynamic dichotomy remains somewhat blurred as implemented in our treatments. Compared to a static-norm design, our dynamic feedback nevertheless clearly puts stronger emphasis on the period-to-period *changes* that are occurring in the reference population (Sparkman and Walton, 2017; Mortensen et al., 2019).

Each feedback letter also includes text on the back, with general information on recycling options in the recipient's municipality as well as some specific tips on how to reduce waste (e.g., by planning food purchases or putting a no-ads sticker on the mailbox). This page did not change over the course of the experiment, although there was some variation across the two municipalities. An example back page (for Varberg) is given in Appendix A.2.

We implement cluster randomization with blocking in both municipalities. The clusters are geographically contiguous groups of addresses that are themselves organized into larger blocks (also contiguous) of exactly three clusters each. Treatment status is perfectly correlated within cluster and each treatment arm is present in all blocks. We use cluster randomization to mitigate potential interference between treatment and control households, which might arise if, for instance, immediate neighbors discuss the letters. Evidence of such across-household spillovers is mixed in previous research (Allcott and Rogers, 2014; Dolan and Metcalfe, 2015; Jaime Torres and Carlsson, 2018). Although randomizing in clusters reduces power to some extent, we prefer to err on the side of caution, not least since our experiments apply norm feedback in a new domain. Furthermore, the use of blocking effectively provides stratification by neighborhood, again increasing estimator precision.[13]

[13]Clusters and blocks were constructed 'by hand' with the explicit objective of sorting similar housing types into the same blocks; for more information on our randomization methodology, see Appendix B.1. Appendix C shows that our sample gives at least 80% power to detect a residual-waste reduction of about 2% (4%) in Varberg (Partille).

VIVAB works to reduce the amount of unsorted waste. With this letter, we want to provide information on the amount of unsorted waste generated by you and your neighbors*. Our garbage trucks weigh each bin during collection. The information in this letter concerns the weight of the unsorted waste. For more information on the contents of the letter, see vivab.info/utskickhushallsavfall

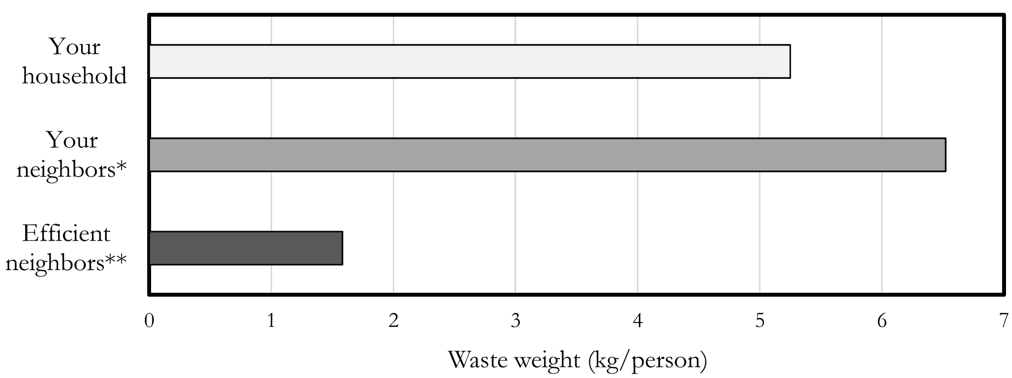
Your own waste and your neighbors' waste, during the period 2019-03-13 to 2019-04-09

During this period, your bin for unsorted waste has been collected three times.

During the latest period, your waste weight was 0,3 kg/person lower than during the preceding four weeks.

Over the same period, 45 percent of your neighbors* have managed to reduce their waste by more than your household.

Waste weights (kg/person) during the period



* The average is based on about 100 households in your neighborhood.
** This 'waste efficient' neighbor represents the average among the 20% of households with the least waste, based on about 100 households in your neighborhood.

Evaluation of your waste weight

EXCELLENT ☺☺ | GOOD ☺ | ROOM FOR IMPROVEMENT

For more information about the letters and how we process your personal data, see vivab.info/utskickhushallsavfall. You can also go there to opt out from receiving additional letters within the project. Please state your service ID: XXXXXX

Figure 2: Example of a dynamic-norm feedback letter

The final feedback letter, sent in late October 2019, informed households that no more letters would be sent and also included a link to an endline survey. The same information was sent without feedback to addresses in the Partille quarterly condition (which had already received feedback one month before) as well as control households, but not to households that had opted out of the study. Survey items included questions regarding knowledge of and attitudes to the project, waste behavior over the preceding months, as well as project-related contacts with other households. Translated versions of the survey are found in Appendix A.3.

4 Data and empirical strategy

Both participating municipalities have had pay-as-you-throw incentives in place prior to the study period and thus weigh all waste bins during collection. The resulting weight records form our main data source and were also used to construct accurate feedback letters throughout the course of our intervention; for information on how this was done, see Appendix B.2.

The raw waste data contain one line per bin-specific collection event, typically including a non-zero weight measured in kilograms. We perform the following operations on these data sets. First, we select all addresses that either received at least one feedback letter, or else are flagged as part of a control cluster. This excludes, for example, a few hundred households flagged as part of a treatment group that did not receive letters due to various exclusion criteria (see Appendix B.2). Our final data set includes 14,935 households in Varberg, and 5,519 in Partille.

Second, the raw waste data involve three waste-bin types: 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.[14] We recode weights associated with the latter pair of waste fractions as a single residual-waste variable. All remaining operations described below are then performed separately for residual and food waste.

Third, collection events may report an associated anomaly whenever, for example, a bin is not placed curbside and thus cannot be collected. For certain such anomaly reports, including when bar codes for bin identification are found to be faulty, we consider stated

[14] As noted in section 2, fractions such as paper and packaging waste are collected through separate channels, in a parallel system based on extended producer responsibility. These fractions are not directly targeted by our intervention, and additionally there is little high-resolution data on their collection. We will return to these points in Section 6.1, where we consider treatment-effect mechanisms.

weights unreliable and recode them as missing. The exact recoding, which is identical to that used when compiling feedback letters, is given in Appendix B.2.

Fourth, most (or, in Varberg, all) households have biweekly collection cycles, with collection from different households roughly evenly staggered across each two-week period. Therefore, we organize our data as an address-by-two-week-period panel. The panel, starting on 19 March, 2018, includes 26 pre-experimental periods ($t \leq 0$), and 18 post-experimental periods ($t \geq 1$). Thus, monthly feedback was received in periods 1, 3, 5, etc; and quarterly feedback was received in periods 1, 7, and 13.[15] To sum the weights within period, we use the following procedure. We first sum all events across individual days, by address and separately for food and residual waste. 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. Then, we sum the resulting day-specific weights across each two-week interval, again by address. Here, any remaining missing values are summed as missing, implying that the two-week period sum will also be missing.

Fifth, the summed weights are then divided by the number of household members as given by register data from the Swedish Tax Authority. For addresses where the tax authority data does not report any household members, values are imputed using the relevant 2019 municipality average for single-family homes from publically available Statistics Sweden data (3.0 persons/household in Partille, 2.7 in Varberg). We are left with two household-level per-capita outcome variables, for residual and food waste, respectively.

Sixth and finally, in accordance with our pre-analysis plan, certain observations and addresses are considered outliers and are dropped from the data. Specifically, we exclude (i) all households with an average residual or food-waste weight above 15 kg/person; (ii) households with >90% missing or zero observations for both residual and food waste, across all periods; and (iii) any single data point with residual or food-waste weight above 50 kg/person. In both municipalities, about 2% of remaining observations are dropped as a result, nearly all of which are excluded due to condition (i) and (ii). Our results are robust to retaining these observations.

Our main regression uses residual waste in kg per person as outcome variable, estimating

$$y_{ijkt} = \lambda_{kt} + \beta_1 T_j^1 + \beta_2 T_j^2 + \theta \bar{y}_i^{PRE} + \boldsymbol{\gamma} \mathbf{X}_i + \epsilon_{ijkt} \tag{1}$$

[15] The periods run from Monday to Sunday at the end of the following week, and do not coincide with the four-week and twelve-week intervals used for feedback purposes, which always run from a Wednesday to a Tuesday. For example, the initial set of monthly letters was compiled on 13 March, 2019 and covered the period 13 February-12 March, which partially overlaps periods −2 to 0.

where i, j, k and t index address, cluster, block, and time, respectively. Since we employ cluster randomization, we consistently cluster robust standard errors at the cluster level (Abadie et al., 2017). λ_{kt} are block by two-week period fixed effects, and $\mathbf{X}_i$ is a set of predetermined address-level controls.

Equation (1) is an ANCOVA regression, replacing address fixed effects with $\bar{y}_i^{PRE}$, the baseline (periods -25 to 0) average of residual-waste weights for household i. ANCOVA can be viewed as an efficient convex combination of difference-in-differences and an *ex-post* comparison of means across treatment arms. It yields weakly higher precision than either component estimator, with efficiency gains compared to difference-in-differences increasing as serial correlation approaches zero (McKenzie, 2012). ANCOVA regressions are run only on post-treatment observations, allowing the treatment t subscript to be dropped. We additionally exclude period 1, when households first received feedback, although we note that results are robust to not doing so. Treatment-group variables T_{jt}^1 and T_{jt}^2 are always equal to zero for control clusters, and are equal to one in associated treated clusters throughout periods 2-18.

5 Results

Figure 3 provides a first look at the experimental results. It tracks average per-person residual-waste weights for each treatment arm and all periods, separately for Varberg (upper panel) and Partille (lower panel). Vertical lines, placed between period 0 and 1, mark the start of treatment.

To the extent that randomization has successfully eliminated average differences between treated and non-treated units, each set of three lines should coincide throughout the pre-treatment period. Reassuringly, this is clearly the case in Varberg despite some rather pronounced seasonal effects.[16] It is not so apparent in Partille, where treatment is randomized over fewer clusters and outcome balance is correspondingly less likely. However, note that, on either side of the dashed vertical line representing the start of treatment, the relative position of each treatment-arm average is roughly constant over time. Thus, while pre-treatment trends do not coincide, they do appear reasonably parallel, suggesting difference-in-differences may be applied as a secondary identification strategy. We return to this point below.

[16] Varberg is a popular domestic summer resort, explaining the peak around periods -18 to -15, at the height of the Swedish summer holidays. Other peaks roughly coincide with national holidays (Easter, Christmas).

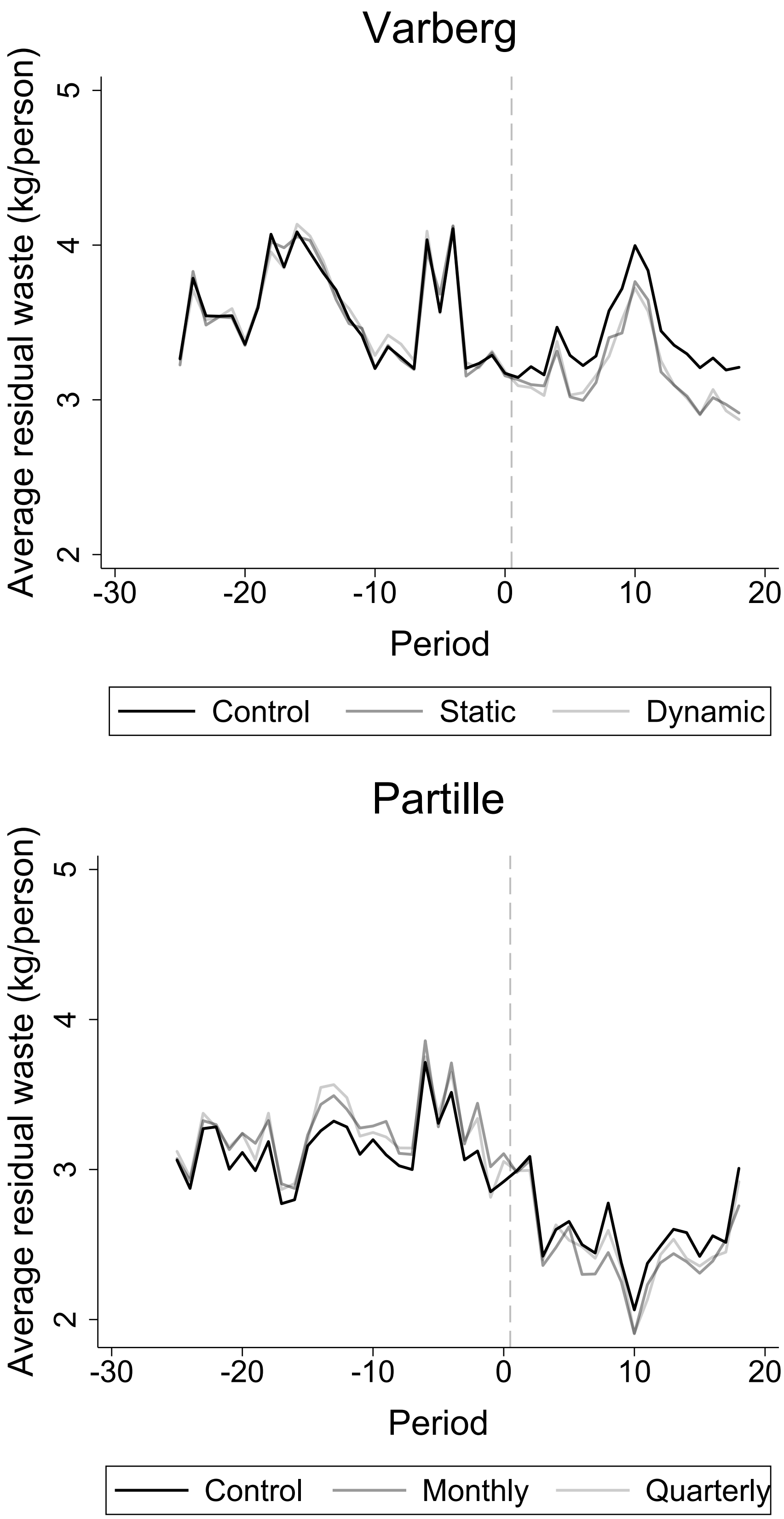


Figure 3: Residual-waste averages by treatment arm and two-week period

Next, to the extent that our interventions are effective, we would expect control and treatment averages to diverge beyond each vertical line. Indeed, it seems that this is happening in both municipalities. In Varberg, from about period 1 onward, control averages are consistently above treated averages, suggesting a residual-waste reduction of about the same magnitude in both the static and dynamic feedback groups. In fact, since pre-treatment trends essentially coincide, these average treatment effects will be roughly equal to the gap between the lines, suggesting a reduction on the order of 0.25 kg/person in both conditions. In relative terms, this is about 7% of the post-treatment control average of roughly 3.40 kg/person. The effect is substantially larger in magnitude than typically found in the literature on Home Energy Reports (e.g., Allcott, 2011, 2015). In Partille, again applying differences-in-differences reasoning, we note that control averages are slightly below monthly and quarterly-group averages up until the start of treatment, and are consistently above thereafter. That pattern again suggests a negative-sign treatment effect, although the magnitude of the effect is less immediately clear than in Varberg.

Table 2 presents ANCOVA regression results. The regression in column 1 corresponds to equation (1) absent covariate vector $\mathbf{X}_i$, confirming a waste reduction of 0.2-0.25 kg/person from both treatments in Varberg. At the bottom of the table, we also report p values for the test that both treatment effects are equal in magnitude; clearly, this null hypothesis cannot be rejected.

Then, in column 2, we add an additional set of household characteristics at baseline, i.e., immediately before the first letter was received. These are: (i) household size, (ii) age of the oldest member of the household (termed head of household), (iii) gender of the oldest member of the household, (iv) whether the household includes at least one child below five years of age; (v) distance, in meters, to the nearest FTI recycling station; and (vi) whether the household's waste collection cycle is two weeks or not. In Partille, about 90% of households in the data have two-week collection cycles, while in Varberg, the figure is exactly 100%, so this covariate is not added there. Our results for Partille are robust to simply dropping those households with collection cycles not equal to two weeks.

With added covariates, treatment-effect estimates are very similar to column 1, indeed slightly larger at 0.25-0.3 kg/person. However, the sample is skewed due to missing covariate data; when we run the regression specification of column 1 on the subsample where covariates are available, we obtain estimates nearly identical to those in column 2. For Partille (columns 3 and 4), ANCOVA estimates for the monthly treatment are similar to either Varberg intervention, both with and without added covariates. Point estimates for quarterly feedback are

	Varberg		Partille	
	(1)	(2)	(3)	(4)
Static (monthly)	-0.218***	-0.263***		
	(0.026)	(0.028)		
Dynamic (monthly)	-0.241***	-0.290***		
	(0.028)	(0.030)		
Monthly (static)			-0.236***	-0.227***
			(0.035)	(0.035)
Quarterly (static)			-0.178***	-0.187***
			(0.038)	(0.037)
Baseline waste average	0.761***	0.755***	0.702***	0.707***
	(0.010)	(0.011)	(0.016)	(0.016)
Household size		-0.013		0.046***
		(0.012)		(0.016)
Age of household head		-0.003**		-0.001
		(0.001)		(0.002)
Male household head		-0.007		0.026
		(0.026)		(0.037)
Child in household		0.196***		0.062
		(0.048)		(0.064)
Recycling-station distance		0.013		-0.243*
		(0.020)		(0.126)
Two-week collection cycle				-0.132
				(0.107)
p value, $\beta_1 = \beta_2$	0.466	0.420	0.126	0.277
Block by period FE	Yes	Yes	Yes	Yes
Observations	250,145	215,943	86,430	83,609
R^2	0.375	0.373	0.393	0.390

Table presents our main ANCOVA regression estimates for average treatment effects on per-person residual waste. Head of household interpreted as oldest member of household. Variable ‘Recycling-station distance’ measured in km. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: *The effect of treatment on per-person residual waste: ANCOVA*

somewhat smaller than for monthly feedback, though the difference is not significant. With the municipality control post-treatment average at about 2.55 kg/person, these estimates correspond to a decrease of 7-9%.

Given the lack of pretreatment outcome balance in Partille, in Table 3 we run regressions like those in column 1 and 3 of Table 2 using difference-in-differences in place of ANCOVA.[17] Note that predetermined controls are invariant within household and thus cannot be used with difference-in-differences. Appendix D.1 supports these regressions by testing the parallel-trend assumption through a series of placebo treatment tests, counterfactually assuming that interventions had begun at various points throughout the actual pre-treatment period. These regressions confirm our earlier conjecture that pre-treatment trends are roughly parallel in both municipalities. Thus, for Partille, the difference-in-difference analysis in Table 3 is our preferred specification. In Varberg, results in Table 3 are nearly identical to the earlier ANCOVA estimates. Point estimates for Partille are slightly larger, at about 0.25 kg/person.

To put these seemingly large reductions into perspective, it is useful to examine what fee increase might have produced similar effects if applied throughout the post-experimental period. In a recent synthetic-control study of Italian unit-based pricing, Bueno and Valente (2019) conclude that a €0.09 per liter volume-based fee reduces unsorted waste by 37.5%, a percentage effect size 4-5 times larger than ours. Assuming a conversion factor of 0.2 kg/liter of residual waste,[18] the Bueno and Valente (2019) fee translates into \$0.54/kg. Thus, as a rough estimate of the equivalent fee increase, we simply divide \$0.54/kg by four or five, yielding an increase of about 32-40% of the current Varberg unit-based fee (\$0.34), and 54-60% of the Partille fee (\$0.2). The effects we observe thus translate into quite large price increases which, strikingly, also exceed the equivalent price increase of 11-20% reported for electricity by Allcott (2011).

All of the above linear-in-parameters regressions do fail to account for two potentially important features of our data. First, about 15% of all weight observations in both municipalities are equal to zero. Although such corner solutions may be due to stringent recycling

[17]All results in both Table 2 and Table 3 are robust to applying a Bonferroni correction for multiple hypothesis testing within regression, adjusting the critical values of the two treatment coefficients as well as that of the $\beta_1 = \beta_2$ test (i.e., $m = 3$).

[18]This conversion factor, also used by Dijkgraaf and Gradus (2004), is additionally supported in the raw waste data. We interpret the upper end of the distribution of weights (in kg) collected from a given type of waste bin as an approximation of its capacity. Since the data also list each bin type's volume, we can calculate kg/liter factors by dividing, e.g., the 99$^{\text{th}}$ percentile of collected weights by the volume. This consistently yields conversion factors around 0.2.

	Varberg	Partille
Static (monthly)	-0.217***	
	(0.028)	
Dynamic (monthly)	-0.246***	
	(0.029)	
Monthly (static)		-0.270***
		(0.037)
Quarterly (static)		-0.212***
		(0.038)
p value, $\beta_1 = \beta_2$	0.371	0.138
Block by period FE	Yes	Yes
Address FE	Yes	Yes
Observations	629,469	217,965
Addresses	14,935	5,519
R^2	0.490	0.518
Within R^2	0.000	0.001

Table presents regression difference-in-differences estimates for average treatment effects on per-person residual waste. Within R^2 relates to remaining variation after absorbing both address and block-by-period fixed effects. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: *The effect of treatment on per-person residual waste: difference-in-differences*

efforts, they are perhaps more likely the result of factors not directly related to waste behavior, such as trips away from the household. Second, the distribution of weights is highly right-skewed; indeed, viewed in a histogram, the empirical distribution essentially decreases monotonically for all positive weights and features a long right tail. This suggests that it may be more appropriate to model the outcome conditional mean as exponential in the covariates.

In Appendix G, therefore, we perform a robustness test accounting for both of these features by estimating the lognormal hurdle model of Cragg (1971). Unlike a standard Type I Tobit model, this approach has the benefit of assuming separate variables and/or coefficients driving corner solutions compared to weight choices conditional on weights being strictly positive. It also assumes an exponential rather than linear model for the interior outcomes. The resulting treatment coefficients (Appendix Table G.1) are comparable but somewhat larger than in Tables 2 and 3, exceeding 0.3 kg/person in some cases and implying reductions of up to 12% compared to the control group. Additionally, treatment-effect differences between monthly and quarterly feedback in Partille are now significant.

Appendix G also presents several other variants of the analysis in Tables 2 and 3: (i) in Table G.2, we do not control for block status in any way, instead including only period fixed effects; (ii) in Table G.3, we collapse the data at the cluster level, using clusters as our unit of analysis and cross-sectional cluster averages of residual waste as outcome variable; (iii) in Table G.4, we pool both municipal data sets and run difference-in-differences regressions that include three treatment variables, where one represents monthly static feedback in both municipalities; and (iv) in Table G.5, we re-run our preferred regressions using residual waste per *household* as outcome variable. All four exercises yield results that confirm those already reported.

Finally, we check for spillovers between treatment and control. In most situations, such interference will reduce any treatment-control difference and bias effect estimates toward zero. For example, members of control households may hear about the feedback letters from receiving neighbors, possibly motivating them to reduce their own waste. Noting that the waste bins are not generally locked in either municipality, *a priori* we also cannot rule out a spillover effect running in the opposite direction: that norm-based feedback induces receiving households to dump some of their waste in a (control) neighbor's bin. This would tend to inflate treatment-control differences, biasing estimates away from zero.

We take a closer look at illicit disposal in section 6.1.1. For now, we note that certain items included in our endline questionnaire are helpful for weighing concerns about dumping; responses are reported in panel A of Table A.1, given in Appendix A.4. Some caution is advised in interpreting the survey responses, given that response rates are generally low: only about 5% of treated households, and 10-18% of control households, participated in the survey. Nevertheless, no more than 10% of control-group respondents claim to be aware of the project, or to have discussed it with others; the overlap is partial, with only 6.6% (3.1%) of control respondents claiming to both be aware of the project and to have discussed it in Varberg (Partille). Among those who have discussed the letters, most did so with someone other than immediate neighbors. Finally, the Partille survey asked respondents whether they thought the letters had 'made any of their neighbors dispose of their waste in an illegal way': strikingly, only about 2.5% of respondents thought so.

Taken together, these results suggest that treatment-control interference is not a major concern. Nevertheless, we are also able to directly check for spillovers between adjacent neighbors in our data. For Partille, we construct a binary indicator for 597 households (32%) that belong to a control cluster while being directly adjacent to at least one treated household. Including the indicator as an additional treatment variable captures any differential effect compared to control households that are not adjacent to treatment, typically because they are in the interior of their cluster. For Varberg, where the large number of households makes such manual coding impractical, we instead interact both treatment variables with a dummy for whether a given block is in a rural area, thus flagging 68 blocks (39.5%).[19] Although the test is admittedly crude, the idea is nevertheless that since houses are spaced further apart in rural areas, spillovers are less likely to occur and thus treatment estimates should exhibit less bias there. Results are given in Table G.6: neither the additional treatment coefficient in Partille, nor the interaction coefficient in Varberg is found to be significant.

6 Extensions

6.1. Mechanisms

What strategies do households use to reduce residual waste in response to feedback? Generally speaking, there are three options available to households. First, as already noted, they may be turning to illicit disposal, i.e., dumping. Second, they might increase their sorting efforts, thus diverting waste from the residual bin to various recyclable fractions. Third, they may reduce the amount of waste generated, for example by buying more packaging-free products. Quantitative analysis is complicated by the fact that, as in the wider economic literature on waste management, little reliable data is available for any of these three waste-reduction categories. Nevertheless, we will discuss each mechanism in turn.

6.1.1. Illicit disposal

As already noted, few respondents in our endline survey (Table A.1) believe illicit disposal is a concern. We are able to complement the survey data by accessing municipal records on dumping incidents related to household waste. These necessarily represent a partial measure, since some types of dumping (e.g., in lakes) are unobservable in the short run. Nevertheless, we would expect any substantial effect on dumping to show up in the records.

In Partille, dumping data are available for 2018 (6 incidents) and the intervention year of 2019 (7 incidents), suggesting no major treatment effect on illicit disposal. Since illicit disposal might also occur across the border of small municipalities like Partille, we also

[19] We visually inspect a map of the municipality to find the blocks corresponding most closely to the set of urban centres (as defined by Statistics Sweden) with at least 300 inhabitants as of 31 December, 2018.

check for dumping incidents in neighboring Härryda municipality, where monthly data on dumping of household waste are available from 2015 through to late 2020. Incident frequency is increasing prior to 2019, so we add a linear time trend in additon to a dummy that equals one from March 2019 onward. The dummy is non-significant ($p = 0.896$), a result which does not change when instead we 'switch off' the dummy after our intervention concluded in October 2019.

In Varberg, municipal records exist from 2015 onward and are disaggregated by waste fraction, allowing us to consider dumping of household waste separately from waste types not targeted by our intervention, such as chemicals, scrap vehicles, and building materials. We do find a spike in dumping incidents related to household waste in 2019 (8 instances, compared to 1-4 during 2015-2018). However, a similar increase appears in 2019 for non-household waste (9 incidents, compared with 2-6 in earlier years), suggesting the variation is unrelated to our experiment. In any case, as in Partille, all incident numbers are clearly extremely small in relation to the number of treated households, so effects on dumping (if any) seem likely to be very minor.

6.1.2. Recycling and prevention

Given that illicit disposal can arguably be ruled out as a mechanism, we now turn to recycling of waste. Recyclables include food, paper, and packaging waste. Starting with food waste, recall that household-specific food weights are available in our main data sets. Figure 4 depicts raw time series for this food-waste variable. Unlike in Figure 3, average pre-treatment weights appear roughly to coincide for all treatment arms in either figure, suggesting ANCOVA regressions may be run in both municipalities. Nevertheless, we also run difference-in-difference regressions in Appendix Table G.7, and supporting placebo regressions in Appendix D.2; the results are very similar to those presented here.

In Figure 4, averages appear to diverge in the post-treatment period, although the effect is much less pronounced than found in Figure 3 for residual waste; note that increased recycling translates into more food waste being collected. ANCOVA regression estimates (Table 4) are consistent with the figure: except for an insignificant and near-zero coefficient for the quarterly treatment, ATEs on food waste cluster around 0.03 kg/person, about one eighth of the reduction in residual waste. We conclude that most of that reduction must be due to other mechanisms.

As for paper and packaging, these waste types are subject to extended producer responsibility regulation and are not collected curbside from single-family homes in Varberg or Partille. Instead, households dispose of them at designated recycling stations. While

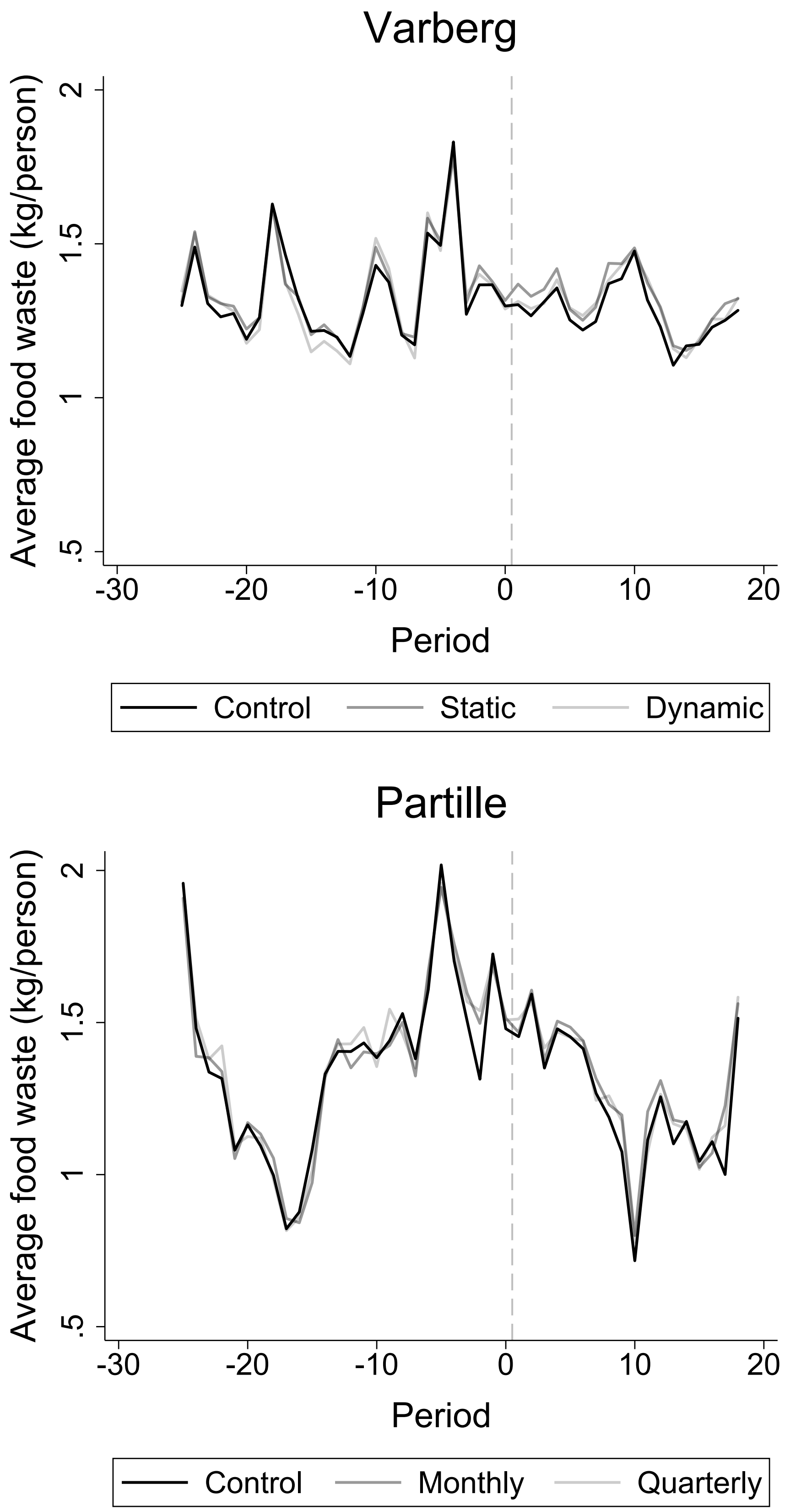


Figure 4: Food-waste averages by treatment arm and two-week period

	Varberg		Partille	
	(1)	(2)	(3)	(4)
Static (monthly)	0.033***	0.021		
	(0.012)	(0.013)		
Dynamic (monthly)	0.027**	0.026*		
	(0.013)	(0.015)		
Monthly (static)			0.037**	0.035**
			(0.016)	(0.016)
Quarterly (static)			0.008	0.002
			(0.016)	(0.016)
Baseline waste average	0.790***	0.788***	0.772***	0.769***
	(0.009)	(0.009)	(0.018)	(0.018)
Household size		-0.003		0.001
		(0.005)		(0.006)
Age of household head		-0.001*		-0.002***
		(0.001)		(0.001)
Male household head		0.024*		-0.003
		(0.013)		(0.016)
Child in household		0.016		0.056**
		(0.018)		(0.025)
Recycling-station distance		0.003		-0.066
		(0.009)		(0.060)
p value, $\beta_1 = \beta_2$	0.668	0.741	0.059	0.023
Block by period FE	Yes	Yes	Yes	Yes
Observations	250,250	216,040	87,969	85,137
R^2	0.443	0.442	0.499	0.496

For both municipalities, table presents ANCOVA regression estimates for average treatment effects on per-person food waste, interpreted as a mechanism for residual-waste reduction. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. Variable 'Two-week collection cycle' not included in the Partille regressions as, for food waste, it equals zero for only about 0.3% of households. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: *The effect of treatment on per-person food waste*

the producer-owned corporation responsible for station maintenance does compile figures for collected paper and packaging by municipality and year, it does so by summing totals across multiple municipalities and then, *ex post*, re-allocating back to each locality on a non-standardized basis. As a result, the impact of any intervention that does not affect re-allocation procedures will be strongly diluted in this data set, and so we do not use it.

An alternative approach to estimating recycling and prevention is to perform a waste composition analysis of residual waste in participating households. The idea is that learning how the post-experimental content of residual waste differs across control and treated households allows us, at least in principle, to infer recycling and prevention behavior for all waste fractions. For example, a lower share of packaging thrown in the residual bin suggests that recycling and/or prevention of packaging waste has increased. We carried out such an analysis in late November 2019, less than a month after the final feedback letter was sent.

The procedure was the following. First, a contractor collected all residual waste generated during a single waste cycle from a (sub)sample of 661 participant households in Partille. The sample was nonrandom but included all single-family households in a particular area within the municipality, and thus was split roughly equally across treatment arms. Two sets of separate but nearly concurrent collection runs were made: one for control households, and one for both treatments. Once all waste had been collected, a random sample of about 500 kilograms (10-20% of collected waste, depending on the group) was made from the waste totals of each collection run. The composition analysis is on that subsample.

While the procedure just described does not permit us to calculate confidence intervals,[20] we present the point estimates, given as weights belonging to each waste type, in Table 5. First, we calculate the weight proportions of each waste fraction. Next, supposing those proportions applied to all participating addresses in Partille, we multiply them either by the control-group post-experimental mean of 2.55 kg/capita (control), or by the same value less the average of the treatment-effect estimates given in column 3 of Table 3, which is 0.241 (pooled treatment). As a result, post-treatment residual waste is decomposed by waste fraction in the rows labeled 'y_{it}'. Finally, we compare the results across control and pooled treatment. Note that we are unable to distinguish recycling from prevention: reductions in

[20]The consultancy in charge of sampling did not separately measure waste from different households. Suppose the subsamples are composed of discrete items (e.g., a metal can or a plastic bottle), with each item a random draw from the corresponding 'supersample' collected by the contractor. Then, given fraction-specific distributions of item weights, one might at least construct weight-proportion confidence intervals based on variation within the supersample. The fraction-specific item weight distributions are unknown, but might be obtained separately and *ex post*, assuming they remain roughly constant across time and space. Unfortunately, plans to do so were derailed by the onset of the COVID-19 pandemic.

	Food	Paper & packaging	Other, avoidable	Other	Sum
Control ($N = 238$)					
Weight (kg)	165.5	200.2	43.1	99.1	507.9
Weight share (%)	32.6	39.4	8.5	19.5	100.0
y_{it} (kg/person)	0.831	1.005	0.216	0.498	2.550
Pooled treatment ($N = 423$)					
Weight (kg)	147.6	192.2	71.2	110.1	521.1
Weight share (%)	28.3	36.9	13.7	21.1	100.0
y_{it} (kg/capita)	0.654	0.852	0.315	0.488	2.309
Reduction: y_{it} (kg/person)	0.177	0.154	-0.099	0.010	0.241

Table presents results from a composition analysis of waste generated by 661 households in Partille. Rows 'y_{it}' estimate weights in kg/person for control and pooled treatment as a whole, multiplying the weight shares of the composition analysis by either the control-group post-treatment mean (2.550 kg/capita), or by the same value less the average of the Partille ATE point estimates in column 3 of Table 3 (0.241). Row 'Reduction' gives the resulting fraction-specific differences across control and pooled treatment.

Table 5: *Results of a waste composition analysis*

Table 5 may be due to either, or both, mechanisms.

Based on the composition analysis, treated households throw 0.177 kg/person less food waste in the unsorted bin than do control households. This is a much larger reduction than the corresponding increase in sorted food waste found in Table 4. Several explanations are possible. First, as noted, some of the reduction computed in Table 5 may be due to prevention and will therefore not show up as increases in sorted food waste. Second, the composition analysis may of course not be representative of Partille as a whole, due to sampling and/or measurement error. Third, the composition analysis treats unopened packaged food as part of the food-waste fraction, but the packaging itself would obviously not add to sorted food waste if such items were correctly source separated. In any case, the reduction in paper and packaging waste is roughly equal to that in food waste, suggesting households respond to treatment by reducing both fractions more or less equally. These reductions are somewhat offset by *increases* in the 'other, avoidable' fraction, including for instance organic (garden) waste, cloth, utensils, etc.

Our endline survey finally provides an alternative source of guidance for judging the relative contribution of recycling and prevention across various fractions. In panel B of Table A.1, we ask control and treated households whether their waste behavior had changed during the intervention period; and if so, what the most substantial change was.[21] Here, increased prevention and/or recycling of food waste appears rare. Indeed, nearly half of households state increased recycling of paper and packaging as their main response to treatment; though stated prevention behavior is also not negligible in our sample.[22]

6.2. Heterogeneous treatment effects

A general finding in the existing literature on norm feedback, whether with respect to water or energy usage, is that treatment effects are largely driven by high users (see e.g., Allcott, 2011; Ferraro and Price, 2013; Jaime Torres and Carlsson, 2018; Andor et al., 2020). Does such a pattern hold within the waste domain as well? To check this, we interact both treatment variables with indicators for each household's position in the distribution of average pre-experimental ($t \leq 0$) waste weights.

For Varberg, the specification used is a modified ANCOVA regression which may be written as

$$y_{ijkt} = q_i^{dec} \times (\lambda_{kt} + \beta_1 T_j^1 + \beta_2 T_j^2) + \theta \bar{y}_i^{PRE} + \epsilon_{ijkt}$$

where q_i^{dec} represents a full set of dummy variables for the baseline waste-weight decile to which address i belongs. Thus, we allow both treatment effects and block-specific time trends to vary by decile.[23] For Partille, recall that our preferred model is instead a difference-in-differences regression where baseline periods remain included in the data. We then run

$$y_{ijkt} = q_i^{dec} \times (\lambda_{kt} + \beta_1 T_j^1 + \beta_2 T_j^2) + \alpha_i + \epsilon_{ijkt}$$

where α_i are address fixed effects, and I is the binary indicator function.

[21]Simple two-proportions tests confirm that changed waste practices are more common among treated households ($p = 0.037$ in Varberg, $p < 0.001$ in Partille). By contrast, chi-square homogeneity tests fail to reject the null hypothesis that the distribution of specific changes made is equal across treatment and control.

[22]For Partille, we are able to support this final point with additional evidence: in July and August 2020, nine months after the last letter was sent, we had research assistants check for 'no-ads stickers' on the mailboxes of all 2,398 single-family homes (split roughly equally across treatment arms) in three specific areas of the municipality. Sticker rates were found to be about 6.4 percentage points higher in the monthly group (two-proportions test: $p = 0.011$), and 4.1 percentage points higher in the quarterly group ($p = 0.101$), compared to control households. We interpret this as evidence of prevention of paper waste, though the amount of waste thus avoided is unknown. Overall long-run effects are discussed in Section 6.4.

[23]Highly similar results are obtained if the decile-specific block-by-time fixed effects are replaced by decile-specific nonparametric time trends along with nonspecific block fixed effects.

For each municipality and treatment, Figure 5 plots the resulting decile treatment effects and associated confidence intervals. Estimates follow the expected pattern, with more pronounced effects among households with high baseline generation. Indeed, reductions in the highest deciles are about 0.5 kg/person (15-20%), twice as large as our estimated ATEs. At the lower end of the baseline distribution, effects instead approach zero, so there is little evidence of a 'boomerang' effect such that low-decile households generate more waste when treated (Schultz et al., 2007).

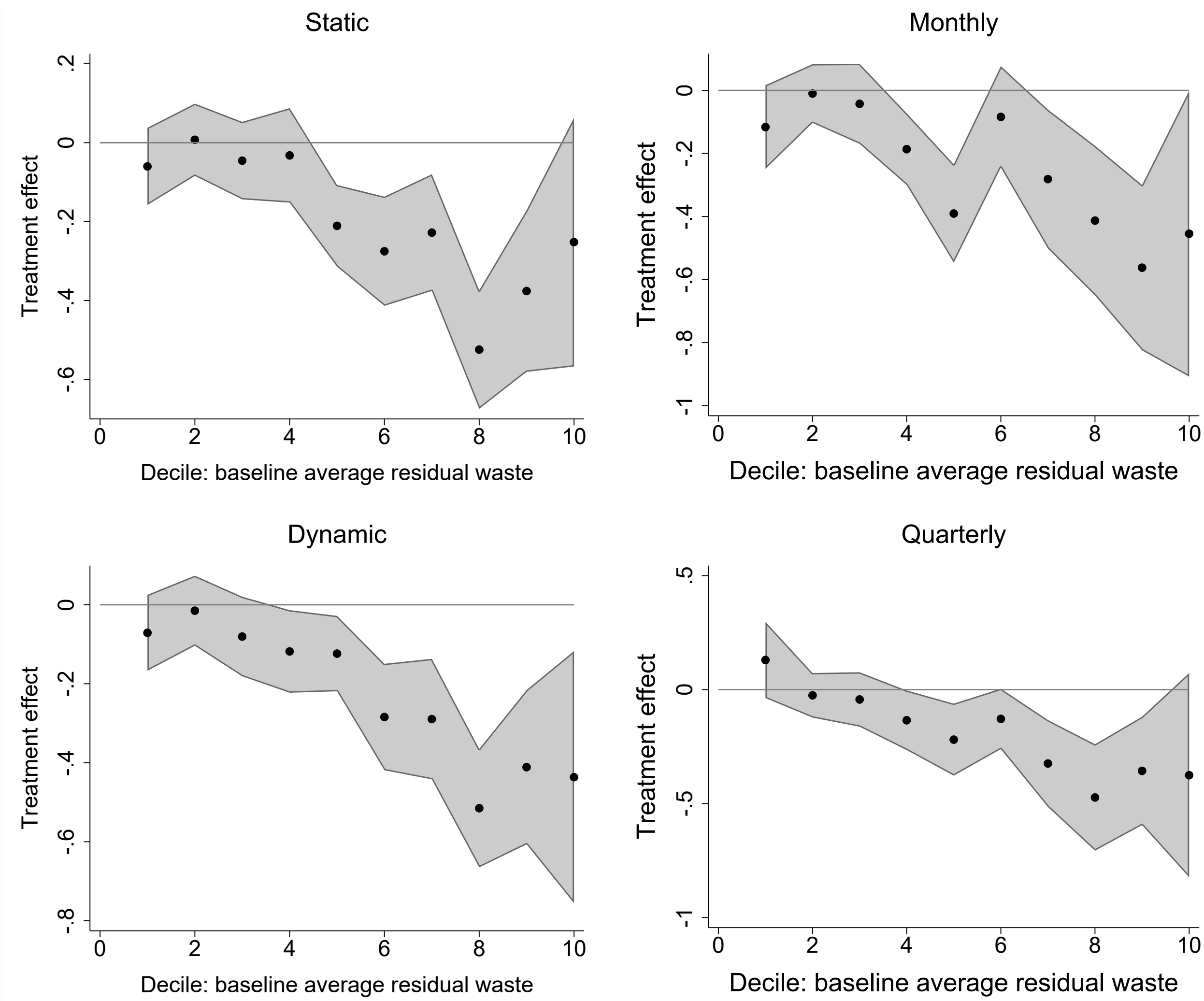
6.3. Effects of injunctive norms: regression discontinuity

A few studies on energy-use feedback (Allcott, 2011; Costa and Kahn, 2013) have also evaluated causal effects specifically of the injunctive norm content included in the letters, i.e., of receiving the rating 'Good' versus 'Room for improvement' (one or no smiley) or 'Great' (one or two smileys). Recall that these ratings are assigned based on the cutoff criterion that a household's weight lies at or below the relevant reference-group average. As a result, (sharp) regression discontinuity (RD) analysis may be applied, comparing households with different ratings in the vicinity of the cutoff.

As a preliminary step, we first compute values of a 'running variable' d_{it} to all residual-waste weights for household i and period t. Here, the running variable is the distance between own-household waste and the relevant cutoff, as shown in the last feedback letter received by i prior to (and including) period t. As a result, cutoffs occur at $d_{it} = 0$. Provided the conditional mean of potential outcomes is smooth at $d_{it} = 0$, observations arbitrarily close to this point on either side will be valid counterfactuals of each other (Imbens and Lemieux, 2008). The implication is that discontinuities in average residual-waste weights across the cutoff may be taken as caused by the switch in ratings.

In Appendix E, we plot residual waste against d_{it} and also estimate discontinuities using local linear regression within a 'bandwidth' around the cutoff. Bandwidths are selected to minimize the mean squared error of the RD estimator, and we additionally use robust bias-corrected inference (Calonico et al., 2014, 2017). These procedures are repeated for both municipalities and separately for the overall and the efficient-neighbor average cutoffs. We also vary whether a single uniform bandwidth is used on either side of the cutoff, or differing bandwidths are allowed; qualitative results are identical across these cases.

In summary, we find all RD estimates to be nonsignificant, except one: in Varberg, there is a positive effect of about 0.2 kg/person from being labeled 'Excellent' rather than 'Good' ($p < 0.01$). This effect, which is about the same magnitude as our main ATEs, stands in contrast to the null results of previous studies such as Allcott (2011). Thus, receiving



(a) Varberg

(b) Partille

Figure 5: Heterogeneous treatment effects by baseline waste generation: point estimates by decile, with 95% confidence intervals

the label 'Excellent' appears to increase subsequent waste generation, a pattern seemingly consistent with the psychological theory of 'moral licensing', where feeling good about oneself perversely causes future bad behavior (Sachdeva et al., 2009; Mazar and Zhong, 2010; Gneezy et al., 2014; Dorner, 2019). However, no such effect arises in Partille, so overall, we would suggest caution in drawing too strong conclusions from the Varberg findings.[24]

[24]Also, note that the replicability of some of the earlier moral-licensing results has been called into question (Blanken et al., 2014; Urban et al., 2019).

6.4. Long-run effects

In early 2021, we obtained an additional batch of waste data from our partner municipalities, allowing us to examine long-run effects in the year after the intervention concluded. The new data set extends to late November 2020 and starts directly at the endpoint of the main data explored thus far, effectively adding post-experimental periods 19-44.

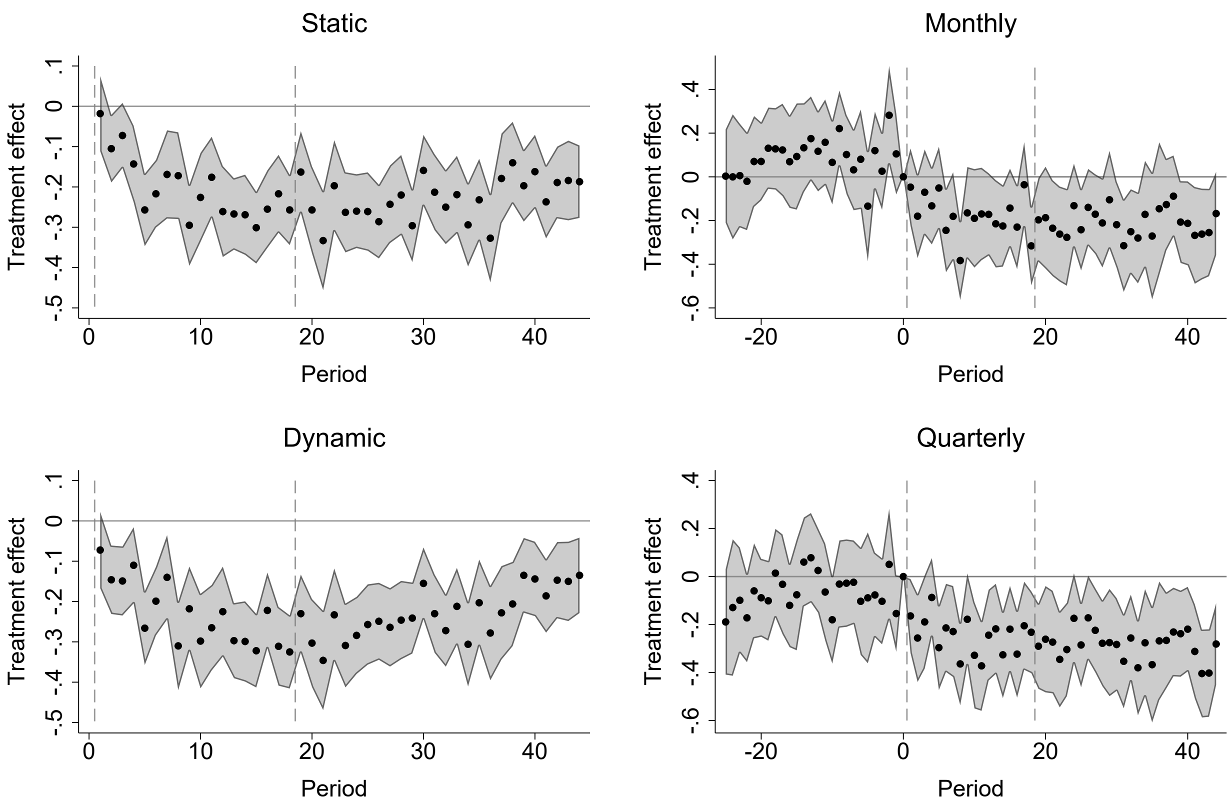
Most HER studies have found that treatment effects are remarkably persistent for both energy (Allcott and Rogers, 2014; Alberts et al., 2016) and water use (Ferraro et al., 2011; Bernedo et al., 2014), often with more than half of the original effect remaining after a year or more. The pattern of immediate change and slow reversion is consistent with some combination of changes in habits and capital stock. For example, Brandon et al. (2017) exploit the fact that in the Opower HER experiments, letters were discontinued upon the sale of a home and not sent to the incoming household. Thus, any residual treatment effect occurring after sale seems likely to reflect changes in physical capital, and Brandon et al. (2017) estimate that this channel accounts for 35-55% of the total effect.

Physical capital plays a different role in the waste domain, however. Waste-related physical capital investments such as in-home recycling bins or 'no-ads' stickers are unlikely to mechanically persist after home sales, limiting the scope for identifying the two channels through home sales. More importantly, effort and physical capital are complements rather than substitutes, as stressed by Vollaard and van Soest (2020). As a result, both short-run and long-run effects will mostly reflect behavior change, with physical capital improvements reduced to a multiplicative effect.

With these points in mind, we examine long-run effects from all four treatments in Figure 6. We run a single regression in each municipality. These interact both relevant treatment variables by all periods except $t = 0$, allowing us to track effect sizes over time.[25] Thus, in the figure, both sets of estimates within a given municipality (e.g., Static, Dynamic) derive from the same regression, and are presented in separate subfigures purely for clarity. The Varberg estimates (panel a), being derived from an ANCOVA specification, use only post-treatment periods, while the Partille difference-in-difference regression (panel b) provides full event studies that include all periods. Dashed grey lines mark the start and end of our original post-treatment period. Both regressions include block-by-period fixed effects but no other covariates (except for baseline averages in Varberg).

Strikingly, we observe quite limited reversion-to-zero effects over the course of the ad-

[25]Note that the regressions include $t = 1$.



(a) Varberg (b) Partille

Figure 6: Long run effects from treatment: period-by-period point estimates and 95% confidence intervals

ditional year. In Partille, such backsliding is practically nonexistent; in Varberg about one quarter (half) of the static (dynamic) norm effect is lost. The patterns are consistent with norm feedback inducing a change in highly sticky, habitual behavior, and are confirmed in additional regressions where treatment status is interacted with a set of longer durations of about six periods each (Appendix Table G.7).

7 Cost, benefits, and effectiveness

In this section, we evaluate the costs and benefits of the feedback interventions. The analysis is necessarily imprecise, because estimates of costs and benefits related to waste and recycling are generally highly uncertain. This is both because valuation of the external benefits is challenging, involving a range of different enviromental impacts, and because estimated costs of waste management vary widely across and within countries. Our aim is to evaluate (i) whether our interventions have increased social welfare and, by extension, (ii) whether expansion of a norm-feedback policy to other Swedish municipalities (with unit-based pricing) would be efficient. We therefore use Swedish data wherever possible.

Figure 7 shows the basic structure of the analysis. The horizontal axis measures the amount of residual waste generated by a (representative) household over the entire course of the intervention. The initial demand curve D_0, assumed linear for simplicity, reflects the household's marginal benefit of generating unsorted waste, or equivalently, marginal costs of reducing it. The presence of unit-based pricing implies pre-existing incentive ϕ.

Our intervention works by shifting the demand curve downward to D_1; the shift is parallel by assumption. Since marginal incentives remain at ϕ, the result is the ATE shown in the figure. This ATE involves several welfare implications. First, internal and external costs associated with the collection and incineration of residual waste drop. Second, internal costs and external benefits associated with the collection, treatment and recycling of food and packaging waste increase. The sum of all these costs and benefits, expressed per kilogram of reduced residual waste, is shown in Figure 7 as the marginal social cost (MSC) of residual waste. Again for simplicity, we draw the line as horizontal. We also place it above marginal incentive ϕ, implying that recycling rates are initially suboptimally low; whether that is the case for our actual setting remains to be seen.

Third, as residual waste drops, households bear higher abatement costs, reflecting some combination of effort and capital investments. (Fee payments are considered transfers and not considered.) Conceptually, these costs are given by a trapezoid area under the demand curve for residual waste, with width equal to the ATE. However, it is not clear whether the right area to use lies below D_0 or below D_1. In a similar welfare analysis, Allcott and Kessler (2019) argue that if a feedback intervention works exclusively by providing information or correcting recipient biases regarding consumption utility, then the area below D_1 is the appropriate measure of household disutility from reduced consumption. If feedback operates exclusively through other channels (e.g., moral pressure), the area below D_0 should be used instead. Following these arguments, we will bound welfare effects by calculating net benefits under either assumption.

In summary, when the D_0 (D_1) cost measure is used, the light (light plus dark) shaded region in Figure 7 represents net social benefits; except for a fourth category of costs not shown in the figure. These are costs directly associated with our intervention, namely project administration costs as well as any household net (dis)utility purely from receiving feedback. Thus, to finally obtain net social benefits, these costs should be subtracted from the shaded

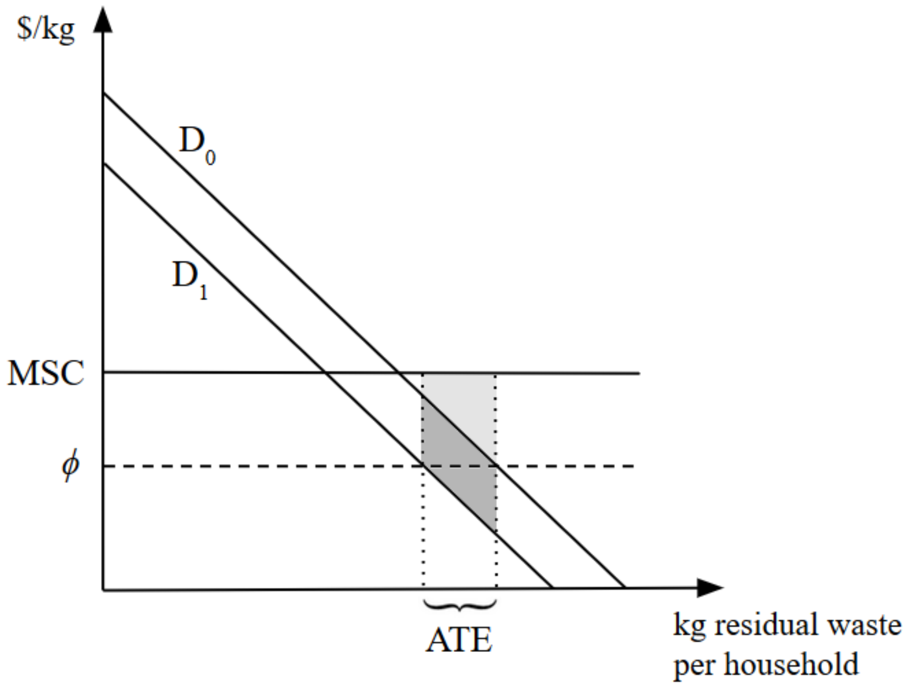


Figure 7: The behavioral and welfare effects of norm-based feedback

area(s).

Table 6 presents our estimates of the components of the analysis, organized by municipality and treatment. To express all costs and benefits in monetary terms, we make use of a range of data sources and assumptions. The first and most complex task is to compute the MSC. Here, we need first to make an assumption regarding what mechanisms underlie any reduction in residual waste. Because our results on the subject (section 6.1.2) are rather inconclusive, we will simply assume that each kilogram of residual waste reduced is matched by a corresponding 750-gram increase in packaging waste, and a 250-gram increase in food waste.[26]

We then obtain estimates of the MSC by combining results from the literature on life cycle analysis (Ambell et al., 2010; Slorach et al., 2019) with environmental shadow prices from Ahlroth and Finnveden (2011). Appendix F describes these calculations in detail. Reflecting a range of estimates from valuation studies, environmental shadow prices may take either a low or a high value. As a result, the MSC may similarly be high or low. The two estimates are very different, mostly because high and low environmental benefits from

[26]Marginal social benefits under high shadow prices (see below) are higher for packaging waste than food waste, so welfare impacts become more favorable as the share attributed to packaging increases.

	Varberg		Partille	
	Static	Dynamic	Monthly	Quarterly
ϕ (\$/kg)	0.34	0.34	0.20	0.20
MSC, low (\$/kg)	-0.41	-0.41	-0.41	-0.41
MSC, high (\$/kg)	0.38	0.38	0.38	0.38
Costs and benefits, high MSC:				
ATE (kg/household, all periods)	26.33	27.57	38.05	27.71
Fee increase yielding ATE (\$/kg)	0.095	0.099	0.164	0.120
Net benefits excl. feedback costs, D_0 (\$/household)	-0.31	-0.39	3.55	3.21
Net benefits excl. feedback costs, D_1 (\$/household)	2.18	2.34	9.80	6.52
Administration costs (\$/household)	5.53	5.53	5.53	1.84
Net benefits (\$/household)				
Low	-5.84	-5.92	-1.97	1.36
High	-3.35	-3.18	4.27	4.68
Allcott and Kessler (2019) 43% rule	0.51	0.79	5.49	6.18
Costs and benefits, high MSC, extrapolated effect:				
Extrapolated ATE (kg/household, all periods)	47.24	33.87	104.82	81.36
Number of post-intervention periods	120	72	120	120
Net benefits (\$/household)				
Low	-5.32	-5.59	4.18	7.20
High	-2.38	-3.07	21.56	17.67
Allcott and Kessler (2019) 43% rule	10.83	7.77	30.35	23.55

Table 6: *Summary of costs and benefits, by treatment*

recycling of packaging differ by a factor of 22. Indeed, the MSC turns out to be *negative* for low shadow prices, in which case the socially optimal recycling rate is clearly zero and any intervention to reduce residual waste will reduce welfare.[27] The high value of the MSC, however, is \$0.38/kg, slightly larger than the pre-existing incentive ϕ in Varberg and almost twice that of Partille. Existing targets and policies for waste strongly suggest that policy makers believe the marginal social benefits of recycling are positive, and we will focus on the high value throughout the remainder of this section.

[27]See Kinnaman (2006) for one cost-benefit analysis with similar results.

Next, to obtain estimates of the ATE shown in Figure 7, we sum the long-run estimates in Appendix Table G.8 across all 44 data periods. For the Varberg static norm, for instance, this total effect equals 26.33 kg/household. As a result, the rectangle below the MSC curve has an area of $0.38 * 26.33 = \$9.89$ for a representative household in this treatment. To calculate household waste-reduction costs (the areas below the demand curves), we again turn to Bueno and Valente (2019) to first calculate the fee increase that, if applied in all post-intervention periods, would produce a total effect equal to 26.33 kg/household. This increase is about 9.5 cents/kg.[28] For a linear demand curve, the area below D_1 in Figure 7 can then be calculated as $26.33 \times 0.34 - (26.33 \times 0.095)/2 = \$7.71$ per household; the area below D_0 is similarly calculated as $26.33 \times 0.34 + (26.33 \times 0.095)/2 = \$10.20$. As a result, net benefits excluding feedback costs equal either $9.89 - 7.71 = \$2.18$ or $9.89 - 10.20 = \$-0.31$, depending on the cost measure used.[29]

Finally, we subtract direct intervention costs as given by the project budget. While our experiments have involved some fixed costs, these are largely due to the initial software coding needed for feedback construction, and are expected to be negligible in any application of our systems to other municipalities. Thus, we subtract only variable operation costs, i.e., due to letter paper, envelopes, printing, packaging, and postal fees. This leaves overall (low or high) net benefits that are either negative or positive, depending on the municipality and to a lesser degree the household cost measure used.

Much of the variation in net benefits is explained by the fact that marginal social costs are already largely internalized in Varberg, with $\phi \approx MSC$. Thus, unlike Partille, even quite low-cost interventions to reduce residual waste further are difficult to justify in Varberg based on economic costs and benefits. The more persistent second-year effects observed in Partille also translate into larger benefit estimates. Notably, the smaller ATE from quarterly feedback is more than compensated for by lower operation costs, suggesting that less frequent feedback can be more efficient.

None of these net-benefit estimates include the direct positive or negative utility impacts households experience purely from receiving feedback. We have no data on such effects,

[28]For Varberg, the 37.5% reduction found by Bueno and Valente (2019) in response to a \$0.54/kg fee corresponds to $0.375 \times 3.4 = 1.275$ kg per person and two-week period. Multiplying by (i) the average number of people per household in Varberg (2.7), and (ii) by the number of post-experimental periods (44), we find that a \$0.54/kg fee may be expected to reduce waste by a total of 151.47 kg/household. The equivalent fee increase is thus $(26.33/151.47) \times 0.54 = \$0.095$/kg. An analogous calculation is performed for other treatments; recall that the control average is 2.55 kg/person in Partille, with 3.0 people per household.

[29]Note that in the latter case, the MSC curve actually intersects D_0, so the lightly shaded area in Figure 7 transforms into two triangles with opposing welfare effects.

and an in-depth examination of the role they play for waste feedback is left for future research. As a speculative exercise, however, we may apply the '43% rule' of Allcott and Kessler (2019). Specifically, they estimate US households' willingness to pay for receiving HER feedback, concluding that total costs to households (including both utility effects from receiving feedback *and* costs from reducing use) amount to 43% of savings from reduced use. Supposing the same figure applies in our setting, then for example, 43% of savings in the Varberg static treatment is $26.33 \times 0.34 \times 0.43 = \$3.85$, implying net benefits of $9.89 - 3.85 - 5.53 = \$0.51$. Table 7 shows that applying the 43% rule generally increases net benefits. Again however, the exercise is only valid to the extent that the ratio of household costs to savings can be translated from the Allcott and Kessler (2019) setting to ours; in particular, the relatively high opt-out rates noted in Section 3 suggest this is a strong assumption.

7.1. Extrapolated total effects

As observed in section 6.4, there are nonzero effects remaining in all treatments at the end of the data period. A reasonable but again somewhat speculative exercise, therefore, is to calculate net benefits for some hypothetical "total" ATE obtained by extrapolating treatment effects until effects have fully dissipated. To perform the extrapolation, we use the following ANCOVA regression specification for Varberg,

$$\begin{aligned} y_{ijkt} = {} & I(1 \leq t \leq 18) \times (\beta_1 T_j^1 + \beta_2 T_j^2) + I(t > 18) \times (\gamma_1 T_j^1 + \gamma_2 T_j^2 + \delta_1 T_j^1 \tau + \delta_2 T_j^1 \tau) \\ & + \lambda_{kt} + \theta \bar{y}_i^{PRE} + \epsilon_{ijkt} \end{aligned}$$

where I is the binary indicator function and τ is a linear time trend. To calculate total effects for treatment l, we sum β_l across initial experimental periods 1-18, and $\gamma_l + \delta_l \times t$ thereafter, until effects have receded to zero. For the static-norm treatment, this happens in period 120, more than four and a half years after the start of treatment.

When we run an analogous difference-in-differences specification for Partille, however, both δ estimates are negative, indicating that effects are without bound, increasing rather than diminishing over time. It seems prudent not to use these coefficients for extrapolation. Nevertheless, to at least provide a rough comparison with results for Varberg, we first sum over the long-run estimates in Appendix Table G.8 and, for periods $t > 18$, simply keep adding the last effect estimated in that table up until period 120.

The results are displayed in the bottom portion of Table 7. Clearly, extrapolation does little to change results in Varberg, but has a large impact on net benefits in Partille, which are now about \$5-20 per household.[30] Applying the 43% rule again increases net benefits even further. The large welfare improvements in Partille arise because the MSC is high compared to ϕ and extrapolated effects do not attenuate over time. In conclusion, net social benefits from feedback are largely determined by the degree to which the financial and external marginal costs of waste management are already internalized through the weight-based fee. When such costs are significantly less than fully internalized, as in Partille, feedback interventions potentially deliver large increases in welfare.

7.2. Cost-effectiveness analysis

Regardless of welfare concerns, it may also be useful to consider the cost-effectiveness of norm feedback compared to other policies that municipalities might use to reducing household waste in line with policy targets. Foremost among these is arguably the replacement of drop-off facilities for packaging waste with curbside collection, thus reducing household effort (and to some extent, monetary) costs associated with recycling. Figures from the Swedish Waste Management Association (2016), a stakeholder organization mainly representing municipalities and public utilities, allows us to calculate the added administrative cost of operating such systems.[31] The report considers two types of curbside collection: optical sorting and four-compartment waste bins. For municipalities with similar characteristics to Varberg and Partille, additional costs of the two system variants are estimated at 130 and 60 SEK per household and year, respectively. Taken over a period of 1.7 years, i.e., the approximate length of our long-run data period, these figures translate into \$26.37 and \$12.17 per household.

By comparison, our total operation cost of providing monthly norm feedback is about \$5.50 per household, and quarterly feedback costs only \$1.84 per household. As a result, curbside collection would need to produce effects on residual waste several times larger than those reported in this paper to compete with feedback interventions. Existing policy evaluations (e.g., Bucciol et al., 2015; Best and Kneip, 2019) suggest effects of about 20%, so the cost-effectiveness of norm feedback seems at the very least on par with alternative approaches for reducing household waste.

[30] Varberg net benefits are nevertheless slightly higher for the extrapolated than the main ATE, despite being negative at the margin in the latter case (footnote 28). This is because, since fee ϕ is assumed to produce a uniform effect across periods, the demand curve in Figure 7 grows increasingly flat as more periods are considered.

[31] These costs do not include fixed costs of switching to the new regime, which typically include purchases of waste bins, vehicles, etc. We expect such costs to be substantial, so the rough analysis presented here is probably (highly) conservative.

8 Concluding remarks

This paper has presented results from two large-scale behavioral interventions providing Swedish households in the municipalities of Varberg and Partille with feedback on how much residual (unsorted) waste they generate compared to neighbors. Depending on treatment and methodology, we estimate immediate average treatment effects (waste reductions) of 7-12%. We also conclude, based on the current state of the art in estimating effects from unit-based pricing of waste, that these reductions correspond to a percentage increase in existing weight-based waste fees of about 30-60%. The results do not appear to be driven by illicit waste disposal by households. Finally, effects are highly persistent, with at least one-third of the immediate reduction remaining one year after treatment was discontinued. Indeed, for the two Partille treatments, effects show no sign of attenuating at all, seemingly increasing rather than diminishing over time.

These results provide strong "proof of concept" that norm-based feedback to households could be a useful non-price tool for reducing waste in line with policy goals. Feedback also appears highly cost-effective compared to curbside packaging collection, arguably the main alternative measure for reducing household residual waste. However, we also find that the net effect of feedback on social welfare strongly depends on the extent to which existing waste fees internalize the marginal social cost of residual waste. In Varberg, where weight-based fees are already close to a (high) estimate of that marginal social cost, feedback is found to have negative net benefits. Moreover, our particular experimental design clearly relies on the pre-existence of unit-based pricing. Thus, an important question to ask is this: Could norm feedback be applied in areas where unit-based pricing systems do not exist, and net benefits are thus likely to be positive and large?

We believe so: a policy maker interested in conducting feedback interventions could, for instance, arrange to lease collection trucks capable of weighing for the duration of the intervention. The resulting temporary weight database could then be used to construct accurate feedback. While such methods are likely to increase intervention costs somewhat, we expect these extra costs to be more than compensated for by the added benefit of reducing waste where no pre-existing marginal incentive exists. It also seems plausible that treatment effects would be larger, since areas without pre-existing incentives are likely to contain more "low-hanging fruits" than already regulated areas like Varberg or Partille.[32] A full exploration

[32] A potential complication related to these points is offered by the recent experiment of Myers and Souza (2020), where norm feedback failed to reduce heating demand among households that do not pay for energy use. The authors suggest that feedback may require pre-existing incentives to be effective. Their setting

of these issues is beyond the scope of this paper, however.

More generally, a limitation of the municipal data used in this paper is that individual households are identifiable only in single-family housing. This problem, which it may be noted does not appear within the electricity domain, needs to be addressed before accurate, household-level feedback can be provided in apartment blocks. On a purely technical level, it may not be insurmountable. For example, utilities could issue household-specific waste bags, similar to bag-based pricing systems, and base feedback on the number of bags thrown per household. However, municipalities may be reluctant to pay the fixed cost of introducing such systems purely to facilitate feedback nudges.

An alternative which remains little explored in the literature on norm feedback is to provide feedback at a more aggregated level. In apartment blocks, it may be feasible to present feedback by building rather than household. We do expect such aggregated feedback to introduce additional problems of coordination and cooperation between households, possibly diluting its effect. On the other hand, existing feedback interventions already presuppose some degree of cooperation between household members (a point also made by Brülisauer et al., 2020), though there is greater scope for communication and observability within than between households. Nevertheless, exploring whether aggregate norm feedback could usefully complement the standard HER design seems a promising avenue for future research.

was distinctive in some respects: a highly environmentally conscious student sample was used, possibly limiting the scope for improvement; and feedback specifically targeted heating rather than overall energy use. Nevertheless, future research should investigate whether their findings apply to other domains.

References

A. Abadie, S. Athey, G.W. Imbens, and J. Wooldridge. When should you adjust standard errors for clustering?, 2017. NBER Working Paper No. 24003.

S. Ahlroth and G. Finnveden. Ecovalue08 — A new valuation set for environmental systems analysis tools. *Journal of Cleaner Production*, 19(17-18):1994–2003, 2011.

G. Alberts, Z. Gurguc, P. Koutroumpis, and R. Martin. Competition and norms: A self-defeating combination? *Energy Policy*, 96:504–523, 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 effect 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. Allers and C. Hoeben. Effects of unit-based garbage pricing: A differences-in-differences approach. *Environmental and Resource Economics*, 45(3):405–428, 2010.

C. Ambell, A. Björklund, and M. Ljunggren Söderman. Potential för ökad materialåtervinning av hushållsavfall och industriavfall, 2010. Report, TRITA-INFRA-FMS 2010:4.

M.A. Andor, A Gerster, J. Peters, and C.M. Schmidt. Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management*, 103, 2020.

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.

C. Berglund. The assessment of households' recycling costs: The role of personal motives. *Ecological Economics*, 56:560–569, 2006.

M. Bernedo, P.J. Ferraro, and M. Price. The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy*, 37(3):437–452, 2014.

H. Best and T. Kneip. Assessing the causal effect of curbside collection on recycling behavior in a non-randomized experiment with self-reported outcomes. *Environmental and Resource Economics*, 72(1203-1223), 2019.

I. Blanken, N. van de Ven, M. Zeelenberg, and M.H.C. Meijers. Three attempts to replicate the moral licensing effect. *Social Psychology*, 45(3):232–238, 2014.

H.S. Bloom. Randomizing groups to evaluate place-based programs. In H.S. Bloom, editor, *Learning More From Social Experiments: Evolving Analytic Approaches*. New York: Russell Sage Foundation, 2005.

A. Brandon, P.J. Ferraro, J.A. List, R.D. Metcalfe, M.K. Price, and F. Rundhammer. Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments, 2017. NBER Working Paper No. 23277.

M. Brülisauer, L. Goette, Z. Jiang, J. Schmitz, and R. Schubert. Appliance-specific feedback and social comparisons: Evidence from a field experiment on energy conservation. 2020.

A. Bucciol, N. Montinari, and M. Piovesan. Do not trash the incentive! Monetary incentives and waste sorting. *Scandinavian Journal of Economics*, 117(4):1204–1229, 2015.

M. Bueno and M. Valente. The effects of pricing waste generation: A synthetic control approach. *Journal of Environmental Economics and Management*, 96:274–285, 2019.

F. Burlig, L. Preonas, and M. Woerman. Panel data and experimental design. *Journal of Development Economics*, 144:102458, 2020.

D.P. Byrne, A. La Nauze, and L.A. Martin. Tell me something I don't already know: Informedness and the impact of information programs. *The Review of Economics and Statistics*, 100(3):510–527, 2018.

S. Calonico, M.D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

S. Calonico, M.D. Cattaneo, and R. Titiunik. Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512):1753–1769, 2015.

S. Calonico, M.D. Cattaneo, M.H. Farrell, and R. Titiunik. rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404, 2017.

S. Carattini, A. Baranzini, and R. Lalive. Is taxing waste a waste of time? Evidence from a Supreme Court decision. *Ecological Economics*, 148:131–151, 2018.

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.

J.G. Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5):829–844, 1971.

M. Czajkowski, N. Hanley, and K. Nyborg. Social norms, morals and self-interest as determinants of pro-environmental behaviors: The case of household recycling. *Environmental and Resource Economics*, 66(4):647–670, 2017.

E. Dijkgraaf and R. Gradus. Environmental activism and dynamics of unit-based pricing systems. *Resource and Energy Economics*, 31(1):13–23, 2009.

E. Dijkgraaf and R. Gradus. An EU recycling target: What does the Dutch evidence tell us? *Environmental and Resource Economics*, 68(3):501–526, 2017.

E. Dijkgraaf and R.H.J.M. Gradus. Cost savings in unit-based pricing of household waste: The case of The Netherlands. *Resource and Energy Economics*, 26(4):353–371, 2004.

P. Dolan and R. Metcalfe. Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation, 2015. Becker Friedman Institute for Research in Economics Working Paper No., 2589269.

Z. Dorner. A behavioral rebound effect. 2019.

C. Ek. Serial-correlation-robust power calculation for the analysis-of-covariance estimator, 2020. Working paper.

C. Ek. A formula for power calculation in cluster-randomized experiments with panel data, 2021. Working paper.

T. Erhardt. Garbage in and garbage out? On waste havens in Switzerland. *Environmental and Resource Economics*, 73(1):251–282, 2019.

I. Ferrara and P. Missios. A cross-country study of household waste prevention and recycling: Assessing the effectiveness of policy instruments. *Land Economics*, 88(4):710–744, 2012.

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.

P.J. Ferraro, J.J. Miranda, and M.K. Price. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–322, 2011.

G. Finnveden, T. Ekvall, Y. Arushanyan, M. Bisaillon, G. Henriksson, U. Gunnarsson Östling, M. Ljunggren Söderman, J. Sahlin, Å. Stenmarck, J. Sundberg, J.-O. Sundqvist, Å. Svenfelt, P. Söderholm, A. Björklund, O. Eriksson, T. Forsfält, and M. Guath. Policy instruments towards a sustainable waste management. *Sustainability*, 5 (3):841–881, 2013.

D. Fullerton and T.C. Kinnaman. Household responses to pricing garbage by the bag. *American Economic Review*, 86(4):971–984, 1996.

U. Gneezy, A. Imas, and K. Madarász. Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11):2645–2658, 2014.

M.H. Heller and A. Vatn. The divisive and disruptive effect of a weight-based waste fee. *Ecological Economics*, 131:275–285, 2017.

S. Holladay, J. LaRiviere, D. Novgorodsky, and M. Price. Prices versus nudges: What matters for search versus purchase of energy investments? *Journal of Public Economics*, 172:151–173, 2019.

J-C. Huang, J.M. Halstead, and S.B. Saunders. Managing municipal solid waste with unit-based pricing: Policy effects and responsiveness to pricing. *Land Economics*, 87(4):645–660, 2011.

G.W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

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*, 92:222–243, 2018.

U. Khan and R. Dhar. Licensing effect in consumer choice. *Journal of Marketing Research*, 43(2):259–266, 2006.

T.C. Kinnaman. Policy watch: Examining the justification for residential recycling. *Journal of Economic Perspectives*, 20(4):219–232, 2006.

T.C. Kinnaman and D. Fullerton. Garbage and recycling with endogenous local policy. *Journal of Urban Economics*, 48(3):419–442, 2000.

G. Kipperberg. A comparison of household recycling behaviors in Norway and the United States. *Environmental and Resource Economics*, 36(2):215–235, 2007.

N. Mazar and C.-B. Zhong. Do green products make us better people? *Psychological Science*, 21(4):494–498, 2010.

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

C.R. Mortensen, R. Neel, R.B. Cialdini, C.M. Jaeger, R.P. Jacobson, and M.M Ringel. Trending norms: A lever for encouraging behaviors performed by the minority. *Social Psychological and Personality Science*, 10(2):201–210, 2019.

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

S. Sachdeva, R. Iliev, and D.L. Medin. Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, 20(4):523–528, 2009.

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.

P.C. Slorach, H.K. Jeswani, R. Cuéllar-Franca, and A. Azapagic. Environmental sustainability of anaerobic digestion of household food waste. *Journal of Environmental Management*, 236:798–814, 2019.

G. Sparkman and G.M. Walton. Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science*, 28(11):1663–1674, 2017.

T. Sterner and H. Bartelings. Household waste management in a Swedish municipality: Determinants of waste disposal, recycling and composting. *Environmental and Resource Economics*, 13(4):473–491, 1999.

Swedish Environmental Protection Agency. Nationell avfallsplan och avfallsförebyggande program 2018-2023: Att göra mer med mindre, 2020. Report. In Swedish.

Swedish Waste Management Association. Beräkning av avfallshanteringskostnader i svenska kommuner, 2016. Report 2016:29.

J. Urban, Š. Bahník, and M.B. Kohlová. Green consumption does not make people cheat: Three attempts to replicate moral licensing effect due to pro-environmental behavior. *Journal of Environmental Psychology*, 63:139–147, 2019.

T. Usui and K. Takeuchi. Evaluating unit-based pricing of residential solid waste: A panel data analysis. *Environmental and Resource Economics*, 58(2):245–271, 2014.

M. Valente. Heterogeneous effects of waste pricing policies, 2020. Working paper.

W.K. Viscusi, J. Huber, and J. Bell. Promoting recycling: Private values, social norms, and economic incentives. *American Economic Review*, 101(3):65–70, 2011.

B. Vollaard and D. van Soest. Breaking habits, 2020. Working paper.

Online Appendices for article “Norm-based feedback on household waste: A large-scale field experiment in two Swedish municipalities”

Appendix A. Experimental materials

A.1 Frequently asked questions

We here present the set of FAQ available to residents of Varberg. Differences between the municipalities were minor, both in terms of what questions were included and in terms of the exact wording of the answers. For example, in Partille we use the term ‘residual waste’ rather than combustible waste, as below. Note that VIVAB is the name of the municipal utility in charge of waste management in Varberg.

VIVAB is currently distributing information on average waste weights to you and other single-family households in your area. By making households aware of how much waste they throw, in kilograms, compared to their negihbors, we are seeking to motivate more people to reduce their total amount of combustible waste.

Varberg municipality uses a weight-based waste fee. This means that, the less waste thrown away by a household, the lower are household costs — and the higher are the environmental gains for a sustainable future in Varberg.

Q. What is the purpose of the project?
A. The purpose is the reduce household combustible waste. Typically, the more waste a household recycles, the less combustible waste will be thrown away. Similar projects on electricity and water in other countries have been shown to reduce environmental impacts.

Q. How will I be affected by the project?
A. If your household is included in the project, you will regularly receive a letter with information on your household’s combustible waste weight. In the letter, you will be compared with the average of about 100 households in your neighborhood, as well as with the average among the 20% of those 100 households that have generated the least combustible waste.

Q. I do not want to participate in the project. What happens once I opt out?
A. You can opt out here [link]. Members of households that choose to leave the project will be immediately excluded from the data set used to compile waste information. However, households that opt out less than a week before a planned letter-distribution date will receive one additional letter.

Q. How long will the project be active?
A. The project will be active until further notice.

Q. Why is the project taking place only in Varberg municipality [i.e., not in Falkenberg, the other municipality serviced by VIVAB]?
A. Varberg municipality has had a weight-based fee since 1994. By comparison, Falkenberg introduced a weight-based fee in 2014.

Q. Is the project financed by municipal taxes or by waste fees?
A. No. The project is funded by the Swedish government. It is not funded by municipal tax funds or by waste fees. We are continually working to improve our information and work processes so that costs are kept as low as possible.

Q. What is the source of the weight information?
A. The garbage trucks of our contractor weigh each waste bin during the collection of household waste. In the project, each summary presented to households is based on those measured weights.

Q. The information concerns household combustible waste. What is that?
A. Combustible waste is what remains after other types of waste, like food waste and packaging, have been separated for use in the production of biogas and biofertilizer as well as for materials recovery.

Q. Will there be any particular consequences as a result of the waste weights presented for my household?
A. No. The project aims only to provide information on your household's combustible waste and allow you to compare your weights with those of households in your neighborhood.

Q. Why do only single-family homes, and not apartment buildings, receive letters?
A. Apartment buildings have common garbage rooms where all households deposit their waste. Since we are unable to separate out the combustible waste generated by each household in apartment buildings, such properties are not included in the project.

Q. Why does my letter have a different design from that of my neighbor?
A. Within the project, we are also exploring various ways of presenting the comparison between households.

Q. My household has not received a letter. Why not?
A. The project has a limited scope and involves only some parts of the municipality. If you have not received a letter, the project does not include your area.

Q. My household shares a waste bin with other households. How have our weights been calculated?
A. We calculate one weight per bin, rather than one weight per household. If a bin is shared by multiple households, however, the weights will be shared by all members of those households. Since, as a result, the same information applies to both households, those you share a bin with have received exactly the same letters as you.

Q. I believe my weights are incorrect. What do I do?
A. The weight information is taken from VIVAB's database of collected weights. These weights are already used as the basis for fee payments. If you suspect that the weights stated in your letter are incorrect, please contact our Customer Service... [contact details]

Q. According to my letter, no weight has been registered. What does that mean?
A. Very rarely, the weight is not registered correctly, for example because an additional bag of waste has been placed next to the bin. In these situations, we are unable to compare your weights with other households.

Q. Is weight information available for food waste as well?
A. At present, the project aims specifically at reducing combustible waste, and only that fraction is included in the letters.

Q. Do you look at the contents of the waste as well?
A. No. Only the weight is being checked, not the contents.

Q. How do you process my personal data?
A. The project uses household-specific information on waste weights. In addition, register data (social security number, name, address) will be processed. These data are handled by a third-party processor, and are used only within this specific project.

Lawful grounds for collection is task in public interest. The project is being operated with the aim of reducing overall waste for environmental benefit and, by extension, a more sustainable society.

Q. My household is not comparable with my neighbors. Why am I still being compared? A. Waste weights are divided by the number of members of each household. We are unable to account for other factors, such as how often household members are present, and this will affect comparability to some degree. Still, we believe households are sufficiently comparable for the information to be relevant.

Q. For the past few weeks, I have not put out my bin for collection. Why am I still being compared?
A. The letters show how many times your bin has been collected within the relevant period, so that the comparison can be put into perspective. Most bins are collected once every two weeks.

Q. Will the information really lead to lower waste? After all, the letters themselves drive an increase in waste.
A. Research on this type of information provision shows that environmental impacts drop significantly. Certainly, the letters themselves involve some additional waste, but we have good reason to believe that the net effect on the environment will be positive.

A.2 Letter back page

Contribute to a better environment: give new life to your waste

Most of our waste can be used to make new products or be recycled as energy. We want to make it easy for you to sort, reuse and recycle as much as possible.

Every sorted piece of packaging makes a difference
Practically any piece of packaging can be used to make new products. If every household recycled one more piece of plastic packaging per month, yearly carbon emissions would be reduced by 3,600 tons, equivalent to the emissions from about 1,200 cars.

Recycling centers in Varberg
The waste that you drop off at one of our three manned recycling centers will be treated in various ways, depending on the type of material. For example, hard plastic is recycled and used to make new plastic products, while pressure-treated wood is considered hazardous waste and is destroyed in incineration plants, where energy is also extracted.

Recycling stations in Varberg
You can drop off old newspapers and packaging of glass, paper, plastic and metal at any of the 25 recycling stations in Varberg. The Packaging and Newspaper Collection, FTI, is responsible for ensuring that the containers are well maintained and emptied regularly. Find a recycling station near you at ftiab.se.

Waste reduction tips from Avfall Sverige

FÖREBYGGANDE

- Avoid having to discard spoiled food by planning your food purchases.
- Stop getting unsolicited advertising mail by putting a 'no ads' sticker on your mailbox.
- Borrow or share seldom-used tools with your friends or neighbors.

ÅTERANVÄNDNING

- Use cloth bags or used paper or plastic bags to pack your grocecies in the supermarket.
- Leave textiles, furniture and building materials at the municipal recycling center. You can also use online second hand sites for buying and selling things.
- Have broken items repaired instead of buying new things.

MATERIALÅTERVINNING

- Leave packaging and newspapers at your nearest recycling station.
- Choose a waste management contract that involves sorting the waste, or compost your organic waste yourself.
- Find more recycling tips at ftiab.se/sortering.html

For more information, see vivab.info

Figure A.1: Example of the back page of a feedback letter (Varberg)

A.3 Endline survey

Here we present the full questionnaire sent to treated households in Varberg; there was only very minor variation in survey design between the two municipalities. The survey was implemented online through the Qualtrics software. Invitation links and QR codes were sent out as part of the final feedback letter (treated households) or as a separate letter (control households, Partille quarterly group). Control households faced a similar but shorter survey to the one below, containing only questions 1, 3, 4 (excl. 4.3), and 5.

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 tried 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! In this questionnaire, we would like you to answer a few questions about the letters and the project.

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 substantial 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: (free text)*
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 100-meter radius)/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 believe the letters lacked? [Answer not required] *Free text*

A.4 Endline survey results

	Varberg		Partille	
	Control	Treatment	Control	Treatment
	$N = 541$	$N = 183$	$N = 322$	$N = 253$
A. Items on spillover effects				
1. *Aware of project?*				
Yes	80 (14.8%)	159 (86.9%)	25 (7.8%)	201 (79.4%)
No	461 (85.2%)	24 (13.1%)	297 (92.2%)	52 (20.6%)
4. *Discussed with others?*				
Yes	55 (10.2%)	128 (70.3%)	22 (6.9%)	128 (50.6%)
No	486 (89.8%)	54 (29.7%)	297 (93.1%)	125 (49.4%)
4.1. *If yes: who with?*				
Neighbor, adjacent	5 (9.3%)	13 (10.2%)	2 (9.1%)	16 (12.9%)
Neighbor, same street	1 (1.9%)	14 (11.0%)	2 (9.1%)	15 (12.1%)
Neighbor, same area	2 (3.7%)	14 (11.0%)	5 (22.7%)	5 (4.0%)
Other	46 (85.2%)	86 (67.7%)	13 (59.1%)	88 (71.0%)
4.2. *If yes: discussed how to reduce?*				
Yes	26 (48.1%)	24 (18.9%)	13 (59.1%)	57 (46.0%)
No	28 (51.9%)	103 (81.1%)	9 (40.9%)	67 (54.0%)
4.3. *If yes: disclosed weights?*				
Yes	N/A	29 (22.8%)	N/A	53 (42.7%)
No	N/A	98 (77.2%)	N/A	71 (57.3%)
Suspect dumping?				
Yes	N/A	N/A	8 (2.5%)	6 (2.4%)
No	N/A	N/A	310 (97.5%)	243 (97.6%)
B. Items on mechanisms				
2. *Letters had an effect?*				
Not at all	N/A	152 (83.1%)	N/A	147 (58.1%)
To some extent	N/A	25 (13.7%)	N/A	79 (31.2%)
To a large extent	N/A	6 (3.3%)	N/A	27 (10.7%)
3. *Changed waste behavior?*				

Yes	50 (9.2%)	27 (14.8%)	25 (7.8%)	60 (23.7%)
No	491 (90.8%)	156 (85.2%)	297 (92.2%)	193 (76.3%)
3.1. *If yes: main change made?*				
More recycling bins in home	8 (16.0%)	6 (22.2%)	1 (4.0%)	4 (6.7%)
More paper/pack. recycled	17 (34.0%)	13 (48.1%)	10 (40.0%)	23 (38.3%)
More food recycled	9 (18.0%)	3 (11.1%)	5 (20.0%)	8 (13.3%)
More paper/pack. prevented	4 (8.0%)	2 (7.4%)	0	3 (5.0%)
More food prevented	2 (4.0%)	1 (3.7%)	2 (8.0%)	12 (20.0%)
Other	10 (20.0%)	2 (7.4%)	7 (28.0%)	10 (16.7%)

Table reports endline survey responses for items related to potential treatment-control interference. Figure N at the top of the table gives total (partial as well as complete) survey responses in each group. 'N/A' means that this particular item was not included in the relevant survey version.

Table A.1: *Endline survey results*

Appendix B. Additional information on design and procedures

B.1 Randomization methodology

The experimental data have a multilevel structure, with clusters nested within blocks. 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 exclude a small number of households (<10%) on the basis of maps and satellite images. In Partille, the excluded group consists mostly of rural households, while in Varberg, it consists mainly of households that are clearly of a different type from their immediate neighbors (e.g., single farms surrounded by residential areas). In both municipalities, we also exclude addresses that are clearly not households (small businesses, etc.).

The remaining households are manually sorted into contiguous blocks of roughly 100 households each. Next, each block is divided as equally as possible into three contiguous clusters of about 30 households each (numbered cluster 1, 2, and 3, typically from northwest to southeast). The clusters are the unit of randomization; treatment status is perfectly correlated within cluster. The blocks thus represent a form of stratification by neighborhood, with all three treatment arms represented within each block. As there are three numbered clusters per block, this creates six possible permutations of treatment arms across the clusters within each block. We use the Stata random number generator to determine which of these six combinations apply within each block.

When constructing the blocks, we take care to ensure that (i) each block consists of similar housing types, and (ii) contains roughly an equal number of households. The point of both rules is to increase precision. Rule (i) does so because treatment effects are estimated across clusters, but within block; thus, between-block (but not within-block) variation will not affect estimator precision. Clearly, it follows that within-block household heterogeneity should be as small as possible. Rule (ii) improves precision because estimator variance is increasing in cluster-size variability, which itself increases more or less mechanically with block variability. The two rules sometimes conflict; since each such situation is unique, we then use discretion to attain reasonable trade-offs.

When dividing blocks into clusters, we again attempt to make splits as equal as possible, similarly to rule (ii) above. In addition, we strive to place cluster borders so as to minimize the number of direct across-border connections between households. This is done to minimize any bias from contamination between treatment and control, which we hypothesize operates between direct neighbors. As a result, households on the same street tend to be assigned to the same cluster, and cluster borders tend to run through back yards and green areas.

An example of the end result of the above methodology is given in Figure B.1. Solid lines are block borders, while dashed lines delineate clusters. Note the ID number in the bottom right of each block. In Varberg, there are 172 blocks, 516 clusters, and a total of 15,723 households. Cluster sizes range between 17 and 45 households per cluster, with an average of 30.47 (SD = 4.63 households). In Partille, we have 55 blocks, 165 clusters, and 5,756 households. Cluster sizes range between 23-43 households per cluster and average 34.88 households per cluster (SD = 3.55 households). Note that our final sample is smaller due for instance to various exclusion criteria applied during the intervention; these are described in Appendix B.2.

Figure B.1: Example of blocks and clusters in northwest Sävedalen, Partille

B.2 Study implementation

Recall that 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, monthly (and, when applicable, quarterly) letters are compiled using scrubbed data from March 2018 (experimental period -25, one year before the start of the intervention) up to the current date. The data set is based on a combination of municipal waste data and register data from the Swedish tax authority. It contains (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, we include no own-household feedback (e.g., own weights) in any letter where either of two conditions apply. First, if no successful collection event has occurred during the latest comparison period, for example due to households never leaving their waste bin out for collection. Second, if at least one collection event is considered unreliable, for example 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, such as when a bin is not placed curbside and thus cannot be collected. Table B.1 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 we disregard the anomaly report.

Furthermore, some households are dropped from the comparison sample during compilation due to noncomparability issues. For example, the data may strongly suggest that a property is currently unoccupied. The exclusion criteria are:

- Varberg. For residual waste, strictly fewer than three two-week periods since March 2018 have one or more successful collection events *as well as* no unreliable collection event

- Partille. For both residual and food waste, no more than 20% of all periods since March 2018 have one or more nonzero-weight successful collection events *as well as* no

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

Table lists possible anomaly incidents, their coding in the raw data, and how each incident is treated during letter compilation.

Table B.1: *Anomaly report coding, Varberg*

unreliable collection event; or, there are collection events of any form in three or fewer periods since March 2018.

Note that these criteria do not exclude outlier weights. Because of the lack of a 20% criterion, fewer households are typically excluded in Varberg than in Partille (recall that the duration since March 2018 is at least one year, that is, at least 26 two-weeks periods). The added criterion for food waste in Partille 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', arising from diligent sorting efforts.

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

Appendix C. Power calculation

In this section, we perform analytical power calculations for ANCOVA based on a 'serial-correlation-robust' procedure that accounts for arbitrary autocorrelation in idiosyncratic errors. The exact procedure is detailed in Ek (2020).[33] It starts from an assumed data generating process (DGP)

$$y_{it} = \lambda_t + \beta T_{it} + v_i + \omega_{it} \tag{C.1}$$

where i and t index unit and time, respectively. The panel is assumed to be balanced. Randomization occurs at the unit level, with a single treatment captured by $T_{it} = 1$ as well as a comparison group for which $T_{it} = 0$. Recall that we randomize by cluster, not unit; we address this discrepancy below. Error terms v_i and ω_{it} are both normally distributed with mean zero, and $Var(v_i) = \sigma_v^2$ while $Var(\omega_{it}) = \sigma_\omega^2$. In addition, ω_{it} is assumed to exhibit arbitrary serial correlation over time.

We now seek to estimate uniform treatment effect β. The experimental data include m pre-treatment periods and r post-treatment periods. Furthermore, there are J experimental units, proportion P of which are treated. Estimation involves running the ANCOVA regression

$$y_{it} = \alpha_t + \tau T_i + \theta \bar{y}_i^{PRE} + \epsilon_{it}$$

on post-treatment observations (so treatment subscript t may be dropped), with robust standard errors clustered by unit. Regression coefficient $\hat{\tau}$ is the treatment estimate.

Given a non-experimental data set, at least $m + r$ periods long and following DGP (C.1) with $\beta = 0$, the power of the above ANCOVA regression estimator may be calculated *ex ante* using the following procedure. First, in each possible continuous panel of length $m + r$ in the non-experimental data, y_{it} is regressed on time and unit fixed effects. From the residuals of this regression, four (co)variance parameters are calculated, stored, and subsequently averaged across all panels. As first shown in Burlig et al. (2020), the experimental ANCOVA estimator variance may be closely approximated as a linear combination of the estimands of the four (averaged) residual-based parameters. In practice, estimates are plugged in in place

[33] Our pre-analysis plan calculated statistical power based on an i.i.d. data generating process, thus ignoring serial correlation. However, Burlig et al. (2020) show that i.i.d. formulas tend to overstate power for long experimental panels such as ours. In line with this point, the results presented here are more conservative than our initial, pre-registered power estimates.

of estimands, yielding an approximate minimum detectable effect

$$MDE^{est} \approx (t^J_{1-\kappa} - t^J_{\alpha/2}) \times \Bigg\{ \frac{1}{P(1-P)J} \times \frac{I}{I-1} \times \Bigg((1-\theta^2)\sigma^2_{\hat{v}}$$
$$+ \left(\frac{m+\theta r}{2m^2r^2}\right)((m+r)(m+\theta r) + (1-\theta)(mr^2 - m^2 r))\sigma^2_{\hat{\omega}}$$
$$+ \left(\frac{m+\theta r}{2mr^2}\right)(m-1)(m+\theta r - (1-\theta)mr)\psi^B_{\hat{\omega}}$$
$$+ \left(\frac{m+\theta r}{2m^2 r}\right)(r-1)(m+\theta r + (1-\theta)mr)\psi^A_{\hat{\omega}} \Bigg) \Bigg\}^{1/2} \tag{C.2}$$

where the quantity in curly braces is the rewritten ANCOVA variance. $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ are suitable critical values of the t distribution with J degrees of freedom; I is the number of units in the non-experimental data set, which may differ from J; and $\sigma^2_{\hat{v}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, and $\psi^A_{\hat{\omega}}$ are (estimands of) residual-based parameters. Finally, in the above equation, ANCOVA weight θ is itself computed as a function of the residual-based parameters, such that

$$\theta = \frac{m(4mr\sigma^2_{\hat{v}} - (m(m-r+2) + r(r-m+2))\sigma^2_{\hat{\omega}})}{2r(2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}})}$$
$$+ \frac{m(-m(m-1)(m-r+2)\psi^B_{\hat{\omega}} - r(r-1)(r-m+2)\psi^A_{\hat{\omega}})}{2r(2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}})} \tag{C.3}$$

We apply these methods to historical data on per-person residual waste, running throughout 2017 and 2018 in Varberg and Partille, respectively. The non-experimental data are organized in the same way and subject to the same exclusion criteria as our actual experimental data set. Thus, they include similar numbers of units (addresses) and span 52 two-week periods. For our calculations, we consistently use $m = 26$, $r = 18$, $P = 0.5$, and significance level $\alpha = 0.05$, as in our main analysis. We also assume statistical power $\kappa = 0.8$.

A few adjustments to the procedure are necessary. As noted, we use clusters of some 30 households as our unit of randomization. Cluster randomization reduces power (Bloom, 2005), so assuming unit randomization will generally imply incorrect results. However, simulations in Ek (2021) suggest ANCOVA estimator precision under cluster randomization is typically quite close to that which applies when the data are first collapsed at the cluster level, and clusters are then interpreted as individual units. In particular, the approximation appears valid when cluster sizes are reasonably balanced, and the ratio of cross-sectional and time-varying variation is not too dissimilar at the cluster and unit levels. For the Varberg data, the coefficient of variation in cluster sizes is rather small, at 0.172. Also, the Stata `mixed` command provides a rough estimate of (i.i.d.) variance components, yielding a cluster-level ratio of cross-sectional to time-varying variation that is 0.633 times that of the corresponding ratio at the unit level.

Given these findings, power calculations on cluster-collapsed data should yield approximately valid results. Thus, we proceed to compute y_{it} as the cross-sectional average per-person waste weight among households in cluster i and period t. In Varberg, $I = 516$ (all clusters in the data) while $J = 344$ (clusters involved in any pairwise treatment-control comparison). We then find that $MDE^{est} = 0.082$ kg/person, about 2.2% of average per-person residual waste in the (non-collapsed) Varberg data.

However, we are unable to repeat the procedure for Partille because the ANCOVA variance constructed from the parameters here turns out to be negative. Notice that for $m = 26$ and $r = 18$, the second and third factors within the bracket in (C.2) may be negative. As a result, it is possible in principle for the reconstructed ANCOVA variance as a whole to sometimes also be negative, given estimation error in the residual-based parameters (Ek, 2020). Obviously, we can also not rule out that model (C.1) is invalid in some respect.

As a second-best approach, we are able to bound the Partille MDE from both sides by observing, first, that ANCOVA is weakly more efficient than difference-in-differences (upper bound). Second, cluster randomization lowers power compared to unit randomization (lower bound). For the upper bound, we run the Burlig et al. (2020) serial-correlation-robust procedure for difference-in-differences (`pc_dd_analytic`) on the cluster-collapsed data ($I = 165$, $J = 110$), yielding $\overline{MDE}^{est} = 0.164$. For the lower bound, we perform the ANCOVA power calculation (C.2) without first collapsing the data at the cluster level ($I = 5,571$, $J = 3,714$). The reconstructed variance is positive in this case and implies $\underline{MDE}^{est} = 0.100$. The resulting MDE range is 3.0-5.0% of average per-person residual waste in Partille. Finally, since m, r, and P do not differ across municipalities, another approach is to combine the residual-based parameter estimates obtained for Varberg with Partille values of I and J: this yields $MDE^{est} = 0.147$, well within the range.

Two features of our data remain unaccounted for. Their effects on power run in opposite directions. First, our design blocks clusters by neighborhood, thus improving precision; second, some 5% of observations are missing, lowering power. We expect the former effect to dominate. As a first approximation of how missing observations affect power, we repeat the Varberg power calculation assuming two fewer time periods (m and r both subtracted by one). This again produces $MDE^{est} = 0.082$ kg/person, very nearly the same value as before. Thus, our MDE estimates seem likely to be conservative overall.

Appendix D. Placebo tests

D.1 Residual waste

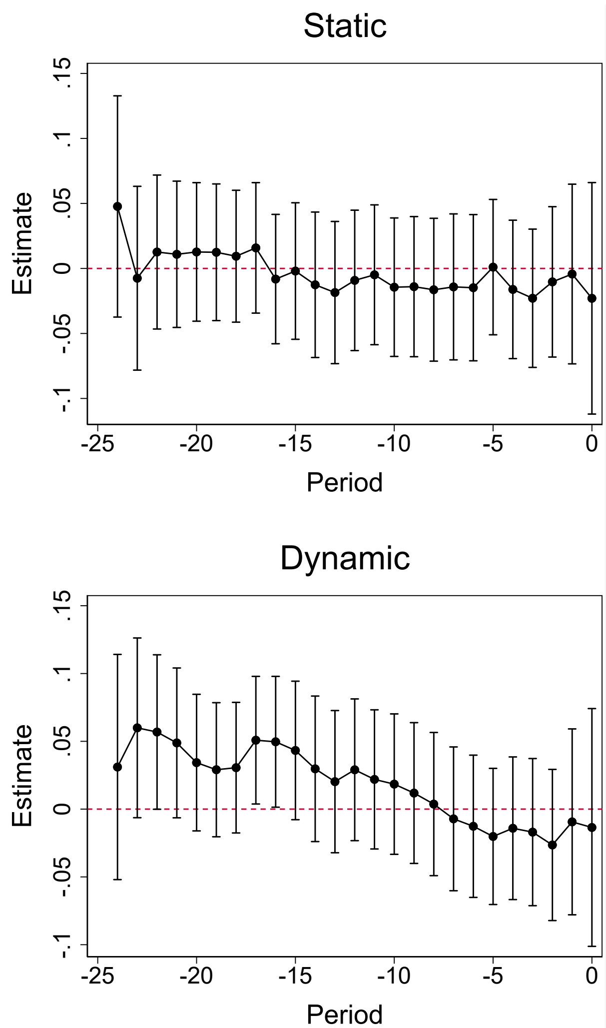
To test the reliability of our difference-in-differences estimates in Table 3 of the main text, we perform a series of placebo regressions where we counterfactually assume that treatment commenced at some point $t_p \leq 0$, i.e., within the actual pre-treatment period. Since each such regression needs to include at least one period before as well as after the onset of the placebo treatment, the first regression has $t_p = -24$, yielding a total of 25 placebo regressions. Each regression is run only on observations with $t \leq 0$ and uses the specification

$$y_{ijkt} = \lambda_{kt} + \delta_i + \beta_1 T_j^1 + \beta_2 T_j^2 + \beta_3 After_t * T_j^1 + \beta_4 After_t * T_j^2 + \epsilon_{ijkt}$$

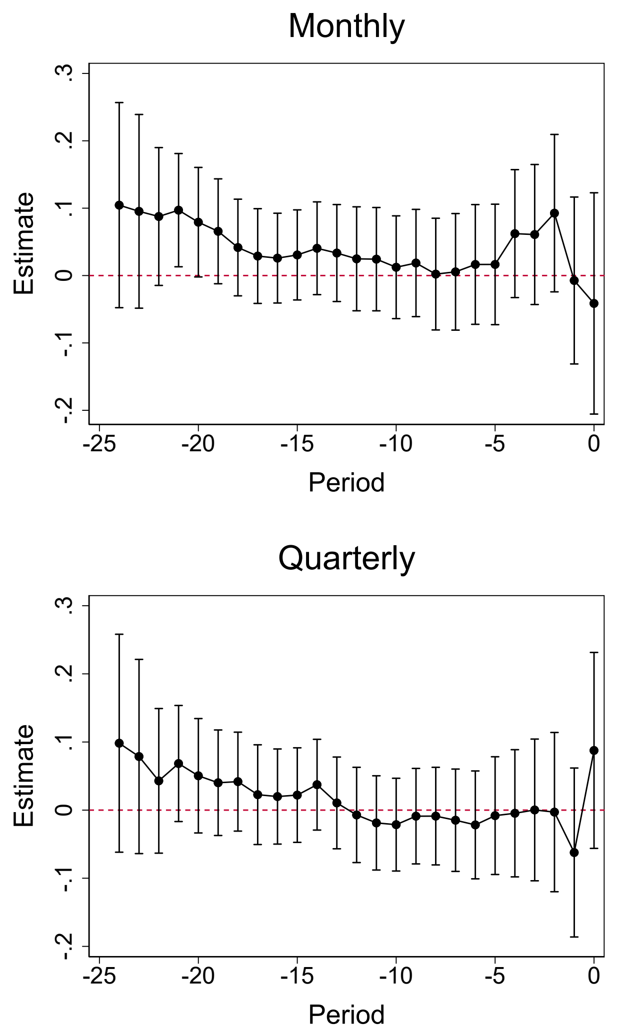
where λ_{kt} is a set of block-by-period fixed effects, and δ_i are unit (address) fixed effects; T_j^1 and T_j^2 are not treatment variables but indicators of (eventual) treatment-group status, such as monthly or quarterly feedback in Partille; and the dummy $After_t$ is equal to one if $t \geq t_p$, and zero otherwise. It follows that placebo treatment effects are captured by interactions $After_t * T_{jt}^1$ and $After_t * T_{jt}^2$. If neither interaction-term parameter estimate is significant, pre-treatment outcomes do not diverge significantly across treatments thoughout $t \geq t_p$ compared to $t < t_p$. Further, if this is found to be the case generally for any t_p, we may conclude that pre-treatment trends are indeed parallel across all three treatment arms.

Figure D.1 reports the results of these tests by plotting, by municipality and treatment arm, the full set of placebo treatment coefficients and associated confidence intervals. Note that we are mainly interested in the placebo coefficients close to the middle of each panel of Figure D.1. For those regressions, both the placebo pre-intervention and the placebo post-intervention periods are relatively long. Thus, short-lived discrepancies between treatment and control will get averaged out, as indeed they are in our main regression. On the sides, by constrast, there is a danger of false positives because the number of pre or post periods is small. For example, in early placebo regressions, coefficients may end up significant because the outcome variable happens to be unbalanced across the few placebo pre-periods; but that significance may well disappear if more placebo pre-periods are added, reducing average treatment-control differences. A similar issue arises for late placebo regressions, where estimates are likewise unreliable due to the limited number of post-intervention periods. Of course, the choice of which placebo regressions to take seriously is arbitrary to an extent, which is why we consistently report all placebo regressions.

In any case, among the 100 interactions estimated for residual waste, two are significant in Varberg (both relate to the dynamic treatment); and only one is significant in Partille ($p = 0.024$). On the whole, we are satisfied that pre-treatment trends are sufficiently parallel, and in particular, that this is so for Partille (panel b).



(a) Varberg

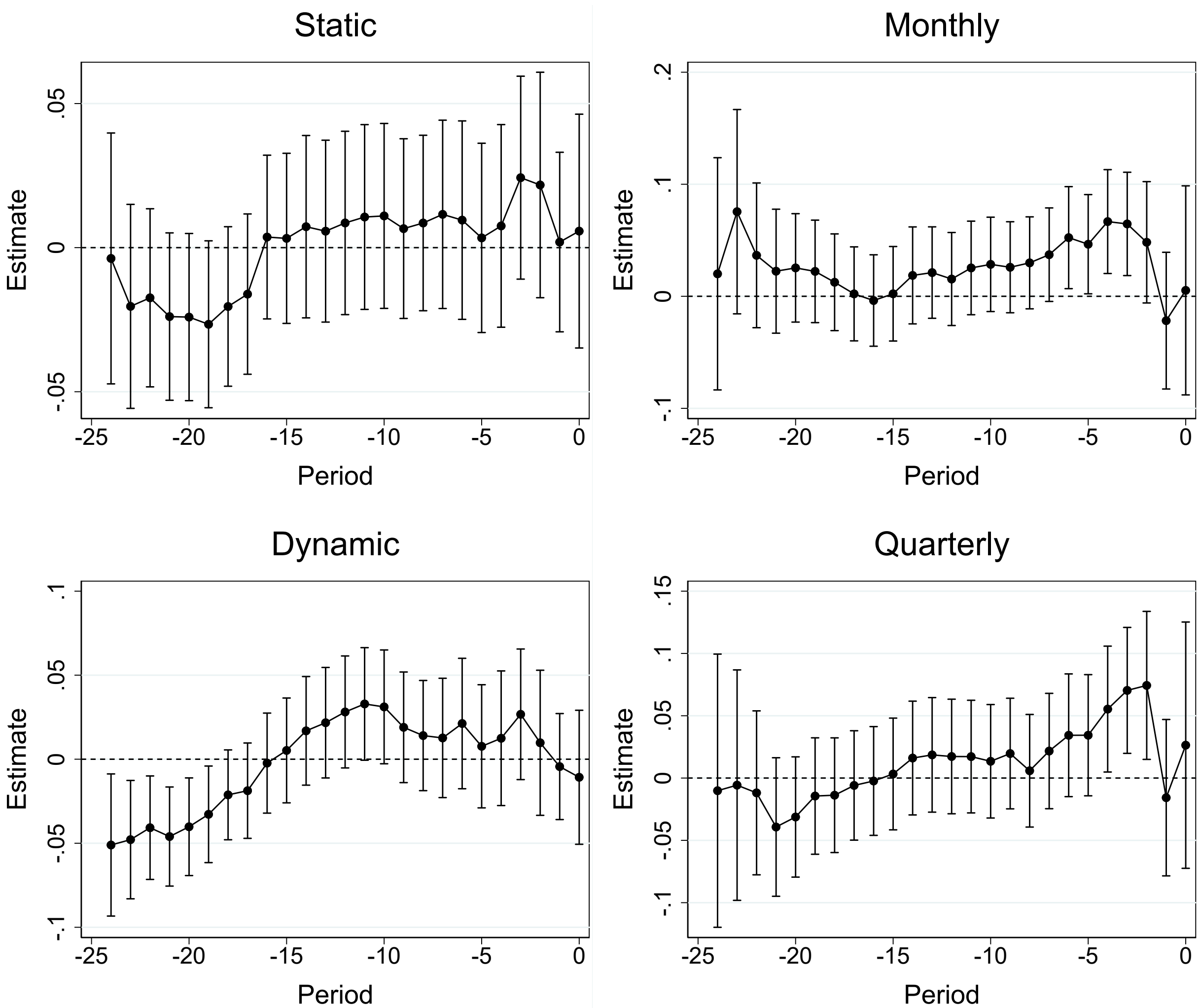


(b) Partille

Figure D.1: Placebo regressions: residual waste

D.2 Food waste

We now repeat the analysis for food waste. In Varberg, we find that a number of early placebo dynamic-norm coefficients are significant; in Partille, the same is true for several late treatment coefficients. For the reasons stated above, however, we are not convinced that these findings should be interpreted as strong evidence that pre-treatment trends are generally non-parallel in our data. As a further illustration of this point, note that averages for the Varberg dynamic treatment in Figure 4 of the main text lie slightly below control averages throughout periods -17 to -12. Indeed, when all periods prior to period -17 are dropped from the pre-experimental data used for placebo regressions, all dynamic-norm treatment coefficients between periods -16 and -9 are found significant at the 5% level. Yet since none of those coefficients are significant in Figure D.2, that significance appears spurious. It seems probable that similar mechanisms are at work in the early and late regressions of Figure D.2.[34]



(a) Varberg (b) Partille

Figure D.2: Placebo regressions: food waste

[34]We have also tried dropping all periods prior to period -18 for the Varberg difference-in-differences regressions in Table G.7. Estimated effects for the dynamic-norm treatment then become somewhat larger, at about 0.035 kg/person.

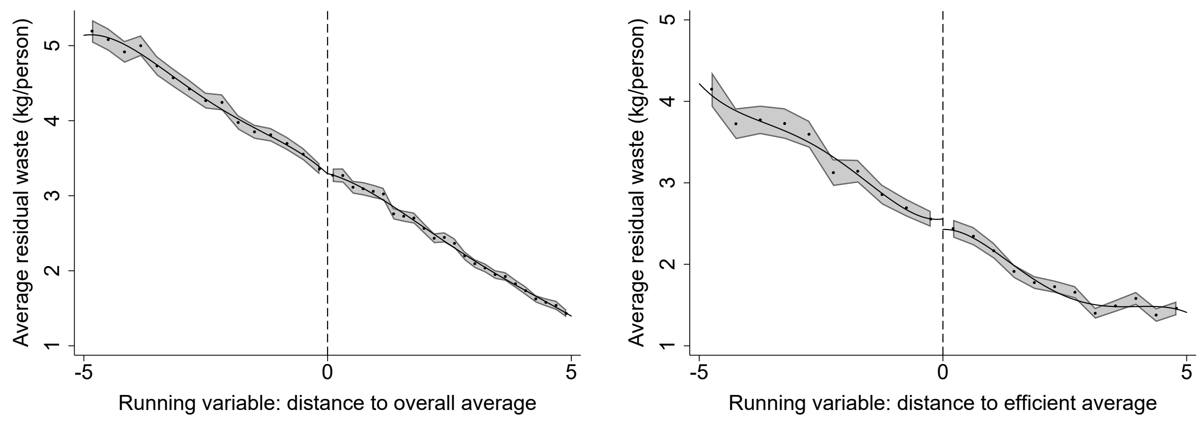
Appendix E. Regression discontinuity analysis

In this section, we perform regression-discontinuity (RD) analyses to estimate the causal effect of a household receiving feedback showing it to lie below (or at) the overall waste average among neighbors ('Good' vs. 'Room for improvement'); or the waste-efficient average ('Excellent' vs. 'Good'). In each case, the running variable d_{it} is the difference shown between the relevant subpopulation average and a household's own waste weight, so the RD treatment applies to all observations with $d_{it} \geq 0$. Specifically, d_{it} uses the distance shown in the last feedback letter received by household i before and including period t.[35]

The key identifying assumption in an RD design is that potential outcomes are smooth, or at least continuous, at the cutoff (Imbens and Lemieux, 2008). Then, units on either side of the cutoff differ on average mostly with respect to discontinuous treatment status; thus, in the limit, any average difference (discontinuity) in observed outcomes at the cutoff may be taken as evidence of a causal effect from treatment. Figure E.1 checks for such discontinuities by plotting the amount of residual waste collected from household i in period t as a function of d_{it}. While later formal RD analyses are based on the full data sets, the figure restricts attention to observations in either municipality with $-5 < d_{it} < 5$. The resulting support of d_{it} on either side of the cutoff is partitioned into a number of equal-sized bins, and waste averages within each bin are represented as solid dots; 95% confidence intervals are also shown in grey. The number and size of the bins are allowed to differ across the cutoff, and are chosen to minimize the mean square error of the bin-average estimators, integrated across the support of d_{it} (Calonico et al., 2015). Finally, to smoothly represent the conditional expectation function, we fit a quartic polynomial, shown as a solid line.

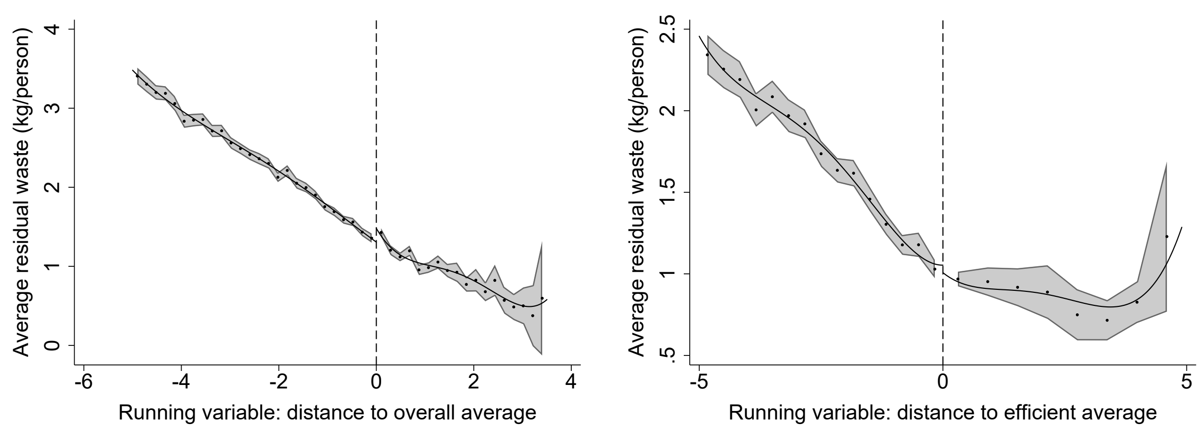
The relationship is clearly decreasing in both municipalities, reflecting serially correlated waste behavior: for instance, households generating large amounts of waste in the comparison period underlying d_{it} (left part of figure) are likely to remain relatively high-waste households in period t. In addition, the lines are very nearly linear. However, there is little evidence of a discontinuity at zero, suggesting no causal effect of a household being given the rating 'Good' rather than 'Room for improvement'. The exercise is then repeated for efficient-mean distance, in Figure E.2. Again, the lines are largely downward sloping and linear, at least

[35] There are no own-household weights coinciding exactly with the overall mean, but a number of observations have exactly $d_{it} = 0$ in relation to the efficient mean. Such households were indeed rated as 'Excellent' in the letters. Also, note that while outliers among own-household weights (>50 kg/person) were not excluded from the feedback letters (Appendix B.2), they are dropped before constructing the pair of running variables.



(a) Varberg (b) Partille

Figure E.1: Residual waste collected as function of running variable: distance to overall neighbor average



(a) Varberg (b) Partille

Figure E.2: Residual waste collected as function of running variable: distance to efficient neighbor average

where there are a reasonable number of observations. In Varberg, there is some indication of a small but precisely identified upward discontinuity at zero, suggesting that households rated as 'Excellent' rather than 'Good' respond by subsequently generating more waste.

To rigorously evaluate potential discontinuities, we perform local linear regression using only those observations that fall within a 'bandwidth' range around the cutoff (Calonico et al., 2017). We consistently use bandwidths that minimize the mean squared error (MSE) of the local linear RD estimator. Moreover, we apply robust bias-corrected inference, first proposed by Calonico et al. (2014). Denoting bandwidths to the left and right of the cutoff as h^- and h^+, respectively, our initial estimation imposes a uniform bandwidth $h = h^- =$

h^+. However, the distribution of waste weights is strongly right-skewed, so our running variables are similarly left-skewed. Thus, as a natural robustness test, we additionally allow for different MSE-optimal bandwidths on either side, so $h^- \neq h^+$.

Results are given in Table E.1. All estimates are nonsignificant except for the Varberg efficient-neighbor average, providing confirmation of the pattern noted in Figure E.2. The coefficient is positive and, at about 0.2 kg/capita, of similar magnitude as the main treatment effect presented in Table 2. No such discontinuity appears in Partille (or the similar analysis of Allcott, 2011), so the evidence is not entirely clear-cut. The Varberg effect does seem consistent with 'moral licensing': receiving explicit praise for one's recycling behavior may lead to a perverse response where such efforts are subsequently reduced. Similar behavioral patterns are documented in, for example, Khan and Dhar (2006), Sachdeva et al. (2009), Mazar and Zhong (2010), and Gneezy et al. (2014); although we also note that the replicability of this literature has recently been called into question (Blanken et al., 2014; Urban et al., 2019).

	Varberg				Partille			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Overall avg.	-0.002	-0.023			-0.054	-0.092		
	(0.051)	(0.046)			(0.076)	(0.074)		
Efficient avg.			0.209***	0.189***			0.011	0.034
			(0.059)	(0.045)			(0.061)	(0.060)
$h^- = h^+$	Yes	No	Yes	No	Yes	No	Yes	No
h^-	2.003	3.999	0.550	2.824	2.713	3.505	3.051	3.445
h^+	2.003	1.563	0.550	0.572	2.713	2.521	3.051	3.777
N^-	20,373	34,125	10,637	53,811	6,881	8,242	15,227	16,699
N^+	29,621	22,366	10,420	10,785	11,311	10,349	7,796	8,134

Table presents robust bias-corrected regression-discontinuity estimates and standard errors, corresponding to the overall average (one smiley; 'Good') or the efficient-neighbor average (two smileys; 'Excellent'). All effects are from left to right, i.e., from higher to lower own-household weights. The (sharp) RD estimates use local linear regression (first-order polynomials) around the relevant cutoff, with mean squared error (MSE) optimal bandwidth(s). h^- and h^+ are optimal bandwidths to the left and right of the cutoff, respectively; we either impose uniform $h^- = h^+$, or allow $h^- \neq h^+$. Similarly, N^- and N^+ are effective observations on either side of the cutoff. Bias estimates are based on second-order polynomials. Variance estimates are clustered at the cluster level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E.1: *Regression discontinuity estimates*

Appendix F. Calculating the marginal social cost of residual waste

In this section, we calculate the marginal social cost (MSC) of residual waste. We do so under the assumption that each kilogram of residual waste causes a corresponding 750-gram increase in packaging waste, and a 250-gram increase in sorted food waste. Thus, the object of interest is the marginal *change* in the sum of internal and external costs as each kilogram of residual waste is traded off against recycled material in this manner. Some of the figures discussed below are direct estimates of cost differences measured across waste fractions. Others are derived (by subtraction) from stand-alone cost estimates for single waste fractions. Wherever possible, Swedish data are used.

Internal costs are financial costs from collection, transport, and disposal of waste, less any revenues from recycled material. External costs arise due to environmental impacts from a given waste fraction; recycling tends to reduce such impacts, yielding a (net) external benefit. In line with Swedish waste policy, we assume that any waste not recycled is incinerated rather than landfilled.

We calculate internal and external cost differentials, compared to residual waste, separately for packaging and food waste. For packaging waste, our figures derive from Ambell et al. (2010), a Swedish-language research paper also summarized in Finnveden et al. (2013). This study compares internal and external costs under two scenarios for 2030: one where Swedish waste management is characterized by "business as usual" (BAU), and one where *all* packaging waste is recycled. The latter scenario thus involves substantial increases in recycled packaging; the first row of Table F.1 reports the magnitudes. For calculating the MSC, we will effectively assume that each 750-gram increase in recycled packaging is allocated across packaging types in proportion to these figures.

The high-recycling scenario in Ambell et al. (2010) is estimated to entail 3800 SEK/-tonne higher internal costs compared to the BAU scenario where more waste is incinerated. After unit conversion, we obtain an estimate of the differential internal cost of packaging recycling, equal to \$0.45/kg.[36] For external costs, we use data on six different environmental impact categories considered by Ambell et al. (2010). Impacts per tonne of packaging type is reproduced in the lower part of Table F.1. Weighting each environmental impact by the top row of the table, we calculate per-tonne impacts from packaging waste as a whole. These weighted averages are given in the last column of the table.

[36] We consistently use the exchange rate 1 USD = 8.38 SEK.

	Paper	Glass	Metal	Plastic	Weighted avg., all
Reduction compared to BAU (tonne)[a]	1,386,000	91,000	264,000	980,000	N/A
Reduction of environmental impacts per tonne recyclable:[b]					
Climate impact (kg CO_2-eq.)	7.99	477	1,956	2,832	1,230
Photochemical oxidants (kg NMVOC-eq.)	2.15	2.00	8.24	6.09	4.15
Terrestrial acidification (kg SO_2-eq.)	2.99	4.58	9.99	4.03	4.10
Freshwater eutrophication (kg P-eq.)	0.065	0.014	0.21	0.008	0.06
Marine eutrophication (kg N-eq.)	1.24	0.52	2.30	1.96	1.58
Depletion of abiotic resources (MJ)	22,047	9,217	29,594	32,909	26,262

[a] Source: Ambell et al. (2010), Table B2. [b] Compared to incineration. Source of impacts: Ambell et al. (2010), Table B3.

Table F.1: *Calculation of MSC of residual waste*

Impact type	Shadow price (SEK/unit), low	Shadow price (SEK/unit), high
Climate impact (kg CO_2-eq.)	0.10	2
Photochemical oxidants (kg NMVOC-eq.)	3	8
Terrestrial acidification (kg SO_2-eq.)	30	30
Freshwater eutrophication (kg P-eq.)	670	670
Marine eutrophication (kg N-eq.)[a]	12	12
Depletion of abiotic resources (MJ)	0.004	0.24
Ozone depletion (g CFC-11-eq.)[a]	1.2	1.2
Freshwater ecotoxicity (kg 1,4-DB-eq.)[a]	60.86	124.37
Marine ecotoxicity (kg 1,4-DB-eq.)[a]	≈ 0	0.606
Terrestrial ecotoxicity (kg 1,4-DB-eq.)[a]	176.47	176.47
Human toxicity (kg 1,4-DB-eq.)	0.004	12

[a] Uses Ecotax02 valuation set. Source: Ahlroth and Finnveden (2011).

Table F.2: *Shadow prices of environmental impacts*

It remains to apply shadow prices to each environmental impact category, so as to obtain external costs per tonne of packaging. We use two sets of prices, both reported in Ahlroth and Finnveden (2011): Ecovalue08 and Ecotax02. The former set includes fewer environmental categories and is used whenever available, with remaining prices taken from the latter set. The relevant shadow prices are reproduced for convenience in Table F.2. Both sets include a high and a low shadow price for most environmental impact categories, reflecting the range of underlying valuation estimates.

External costs are then computed in Table F.3. The first column simply repeats each environmental impact per tonne of packaging waste. Next, the second and third columns monetize all impact categories (as SEK/tonne of packaging) using low and high environmental shadow prices, respectively. Finally, the bottom of the table reports summed external costs per unit of packaging: \$0.05 for low shadow prices, and \$1.07 for high prices.

Ambell et al. (2010) does not consider increased recycling of food waste. To obtain internal costs for that waste fraction, we turn to a report by the Swedish Waste Management Association (2016), a stakeholder organization mainly representing municipalities and public utilities. The report calculates per-tonne financial costs of waste management, net of revenue from recycled materials, separately for food and residual waste. This is done for a range of systems and municipality types. The values most applicable to Varberg and Partille (report appendix 6, bottom table, column 2) are for a municipality without curbside collection, with above-median population in urban centres (Varberg is in the 14$^{\text{th}}$ percentile nationally, Partille in 20$^{\text{th}}$ percentile) and above-median share of apartment housing (Varberg is in the 29$^{\text{th}}$ percentile, Partille in 15$^{\text{th}}$ percentile). While considerable variation likely remains even within this group of municipalities, we then estimate the food-residual cost differential at 3,500 SEK/ton, or \$0.42/kg of food waste.

We have found no estimate of the external benefits of food-waste recycling (i.e., of anaerobic digestion compared to incineration) specific to Sweden. We therefore use a life cycle analysis performed on UK data (Slorach et al., 2019) to calculate these benefits. The authors consider a very wide range of environmental impact categories; we limit the analysis to those impacts for which shadow prices are available in Ahlroth and Finnveden (2011). Differential impacts per tonne of food waste, calculated as the difference between digestion and incineration in Slorach et al. (2019), are reproduced in Table F.3, column 4. The final two columns then monetize each impact using low and high shadow prices, respectively. Note that digestion implies *larger* environmental impacts than incineration along some dimensions: in particular, digestion causes greater terrestrial acidification due to emissions of

	Impact per tonne packaging	Packaging external benefits, low	Packaging external benefits, high	Impact per tonne food[a]	Food external benefits, low	Food external benefits, high
Climate impact (kg CO_2-eq.)	1,230	122.99	2,460	29	2.9	58
Photochemical oxidants (kg NMVOC-eq.)	4.15	12.46	33.22	0.69	2.07	5.52
Terrestrial acidification (kg SO_2-eq.)	4.10	122.93	122.93	-7.2	-216	-216
Freshwater eutrophication (kg P-eq.)	0.06	38.27	38.27	0.02	11.46	11.46
Marine eutrophication (kg N-eq.)	1.58	18.94	18.94	-0.6	-7.2	-7.2
Depletion of abiotic resources (MJ)	26,262	105.05	6,303	1,050	4.2	252
Ozone depletion (g CFC-11-eq.)				0.005	0.006	0.006
Freshwater ecotoxicity (kg 1,4-DB-eq.)				1.82	110.77	226.35
Marine ecotoxicity (kg 1,4-DB-eq.)				1.6	0	0.97
Terrestrial ecotoxicity (kg 1,4-DB-eq.)				0.006	1.09	1.09
Human toxicity (kg 1,4-DB-eq.)				21.6	0.09	259.2
Total (SEK/tonne recyclable)		420	8,976		-91	591
Total ($/kg recyclable)		0.05	1.07		-0.01	0.07

[a] Source: Slorach et al. (2019), Figure 4; these impacts are for anaerobic digestion in comparison with incineration. For each environmental category, units given in parentheses pertain to (reduced) impacts, not to external benefits, which are in SEK/tonne of recyclable waste and are summed on row "Total (SEK/tonne recyclable)". Note that anaerobic digestion has a larger environmental impact than incineration along certain dimensions, in which case impact and cost figures will be negative.

Table F.3: *Calculation of MSC of residual waste* ammonia when applying digestate to farmland. As a result, external benefits of food-waste recycling are a *negative* -\$0.01/kg when low shadow prices are used. For high shadow prices, external benefits are \$0.07/kg.

With these results, we are finally in a position to calculate the MSC per kilogram of residual waste. Under low shadow prices, this is $0.75 \times (0.05 - 0.45) + 0.25 \times (-0.01 - 0.42) = -\$0.41$. High shadow prices instead imply $0.75 \times (1.07 - 0.45) + 0.25 \times (0.07 - 0.42) = \$0.38$. These are the values used to complete our cost-benefit analysis in the main text.

Appendix G. Additional regression tables

	Varberg		Partille
	(1)	(2)	(3)
Static (monthly)	-0.277***	-0.309***	
	(0.032)	(0.031)	
Dynamic (monthly)	-0.272***	-0.324***	
	(0.030)	(0.032)	
Monthly (static)			-0.314***
			(0.057)
Quarterly (static)			-0.247***
			(0.054)
Baseline log waste average	0.845***	0.838***	
	(0.010)	(0.011)	
Household size		-0.075***	
		(0.018)	
Age of household head		-0.005***	
		(0.001)	
Male household head		0.038	
		(0.032)	
Child in household		0.452***	
		(0.067)	
Recycling-station distance		0.006	
		(0.024)	
p value, $\beta_1 = \beta_2$	0.791	0.944	0.004
Block by period FE	Yes	Yes	Yes
Address dummies	No	No	Yes
Observations	250,146	215,943	217,966
Pseudo R^2	0.158	0.172	0.313

Table presents Cragg (1971) lognormal hurdle estimation of average treatment effects on log per-person residual waste. We report unconditional marginal effects on the observed waste-weight variable, y_{it}. In all columns, block-by-period FE are included as dummies in the second-step equation, but we include only block and period dummies (no interactions) in the first step estimating the probability of nonzero weights. In the Partille regressions (columns 3 and 4), address dummies ('fixed effects') are included in the second step to mimic our preferred difference-in-differences approach; note that the parameters in the second step are equivalent to those of a log-linear OLS regression on observations where $y_{it} > 0$. We do not include address dummies in the first (probit) step, because this would lead to an incidental-parameters problem where the number of parameters grows with the number of observations and estimates are thus inconsistent. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. In the hurdle regression, we cluster robust standard errors at the cluster level. Standard errors for the marginal effects are calculated using the `unconditional` option and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.1: *Estimating treatment effects using a lognormal hurdle model*

	Varberg		Partille
	(1)	(2)	(3)
Static (monthly)	-0.220***	-0.273***	
	(0.033)	(0.035)	
Dynamic (monthly)	-0.241***	-0.288***	
	(0.040)	(0.041)	
Monthly (static)			-0.258**
			(0.113)
Quarterly (static)			-0.205*
			(0.115)
Baseline waste average	0.760***	0.757***	
	(0.010)	(0.011)	
Household size		-0.012	
		(0.012)	
Age of household head		-0.002*	
		(0.001)	
Male household head		-0.010	
		(0.026)	
Child in household		0.196***	
		(0.047)	
Recycling-station distance		-0.010*	
		(0.006)	
p value, $\beta_1 = \beta_2$	0.599	0.720	0.627
Period FE	Yes	Yes	Yes
Address FE	No	No	Yes
Observations	250,146	215,943	217,966
R^2	0.353	0.352	0.001 (within)

Table presents ANCOVA (Varberg) and difference-in-differences (Partille) estimates for average treatment effects on per-person residual waste. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.2: *The effect of treatment on per-person residual waste: period FE only*

	Varberg		Partille
	(1)	(2)	(3)
Static (monthly)	-0.221***	-0.217***	
	(0.026)	(0.026)	
Dynamic (monthly)	-0.233***	-0.233***	
	(0.027)	(0.027)	
Monthly (static)			-0.280**
			(0.115)
Quarterly (static)			-0.219*
			(0.120)
Baseline waste average	0.769***	0.760***	
	(0.030)	(0.029)	
Household size		0.006	
		(0.072)	
Age of household head		-0.003	
		(0.005)	
Male household head		0.077	
		(0.120)	
Child in household		0.078	
		(0.305)	
Recycling-station distance		0.042*	
		(0.023)	
p value, $\beta_1 = \beta_2$	0.669	0.579	0.589
Block and period FE	Yes	Yes	Yes
Address FE	No	No	Yes
Observations	8,735	8,734	7,041
R^2	0.422	0.424	0.006 (within)

Table presents ANCOVA (Varberg) and difference-in-differences (Partille) estimates for average treatment effects on cluster-collapsed per-person residual waste. We collapse the data by calculating cross-sectional cluster averages of (nonmissing) values of all variables. Note that, since the number of addresses that we average over within each cluster may vary across time periods, these quantities will differ somewhat from the overall within-cluster average of each cluster throughout the pre-intervention period. All regressions in the table include only block and period FE; adding block-by-period interactions as well would imply very few observations per estimated parameter. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.3: *The effect of treatment on per-person residual waste: cluster-collapsed data*

	Partille and Varberg
Static, monthly	-0.231***
	(0.023)
Dynamic, monthly	-0.253***
	(0.028)
Static, quarterly	-0.192***
	(0.036)
p value, $\beta_1 = \beta_2 = \beta_3$	0.375
Block by period FE	Yes
Address FE	Yes
Observations	847,434
Within R^2	0.001

Table pools both municipalities into a single large data set and presents corresponding difference-in-differences estimates for average treatment effects on per-person residual waste. The p value reported at the bottom of each column relates to an F test of the null hypothesis that all three treatment coefficients are equal. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.4: *The effect of treatment on per-person residual waste: pooled data*

	Varberg		Partille
	(1)	(2)	(3)
Static (monthly)	-0.512***	-0.599***	
	(0.061)	(0.065)	
Dynamic (monthly)	-0.508***	-0.614***	
	(0.066)	(0.071)	
Monthly (static)			-0.752***
			(0.114)
Quarterly (static)			-0.659***
			(0.114)
Baseline waste average (kg/household)	0.836***	0.808***	
	(0.007)	(0.009)	
Household size		0.130***	
		(0.033)	
Age of household head		-0.017***	
		(0.003)	
Male household head		-0.068	
		(0.062)	
Child in household		0.607***	
		(0.166)	
Recycling-station distance		0.076*	
		(0.041)	
p value, $\beta_1 = \beta_2$	0.954	0.840	0.451
Period FE	Yes	Yes	Yes
Address FE	No	No	Yes
Observations	250,145	215,943	217,965
R^2	0.502	0.517	0.001 (within)

Table presents ANCOVA (Varberg) and difference-in-differences (Partille) estimates for average treatment effects on residual waste per household (address). Baseline averages are also in kg per household. For comparison, the post-treatment control average of this variable is 8.83 kg/-household in Varberg and 7.58 kg/household in Partille. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.5: *The effect of treatment on residual waste per household*

	Varberg		Partille
	(1)	(2)	(3)
Static (monthly)	-0.196***	-0.236***	
	(0.035)	(0.036)	
Static (monthly) × Rural	-0.061	-0.079	
	(0.052)	(0.056)	
Dynamic (monthly)	-0.260***	-0.308***	
	(0.038)	(0.040)	
Dynamic (monthly) × Rural	0.055	0.051	
	(0.054)	(0.058)	
Rural	-0.392	-0.353	
	(0.352)	(0.327)	
Monthly (static)			-0.279***
			(0.042)
Quarterly (static)			-0.221***
			(0.043)
Adjacent to control			-0.029
			(0.069)
Baseline waste average	0.761***	0.755***	
	(0.010)	(0.011)	
Household size		-0.013	
		(0.012)	
Age of household head		-0.003**	
		(0.001)	
Male household head		-0.008	
		(0.026)	
Child in household		0.196***	
		(0.048)	
Recycling-station distance		0.013	
		(0.020)	
Block and period FE	Yes	Yes	Yes
Address FE	No	No	Yes
Observations	250,146	215,943	217,965
R^2	0.375	0.373	0.001 (within)

Table checks for spillovers between treatment and control groups in the context of ANCOVA (Varberg) and difference-in-differences (Partille) regressions. 'Rural' and 'Adjacent to control' are both dummy variables. Rural indicates whether the block that a household belongs to is urban (blocks 1-82) or rural (blocks 83-172). Adjacent to control is equal to one for control households directly adjacent to a treated cluster; 597 households, 32% of the control group, were flagged in this fashion. Head of household interpreted as oldest member of household. Variable 'Recycling-station distance' measured in km. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.6: *Spillovers between treated and control clusters*

	Varberg	Partille
Static (monthly)	0.030**	
	(0.013)	
Dynamic (monthly)	0.029*	
	(0.015)	
Monthly (static)		0.036**
		(0.018)
Quarterly (static)		0.003
		(0.018)
p value, $\beta_1 = \beta_2$	0.919	0.034
Block by period FE	Yes	Yes
Address FE	Yes	Yes
Observations	629,653	225,004
Addresses	14,935	5,519
R^2	0.519	0.568
Within R^2	0.000	0.000

For both municipalities, table presents regression difference-in-differences estimates for average treatment effects on per-person food waste, interpreted as a mechanism for residual-waste reduction. Within R^2 relates to remaining variation after absorbing both address and block-by-period fixed effects. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.7: *The effect of treatment on per-person food waste: difference-in-differences*

	Varberg		Partille
	(1)	(2)	(3)
ATE, Treatment 1:			
$1 \leq t \leq 6$	-0.135***	-0.172***	-0.203***
	(0.027)	(0.029)	(0.039)
$7 \leq t \leq 12$	-0.216***	-0.267***	-0.295***
	(0.035)	(0.036)	(0.042)
$13 \leq t \leq 18$	-0.261***	-0.313***	-0.279***
	(0.032)	(0.034)	(0.046)
$19 \leq t \leq 24$	-0.244***	-0.297***	-0.299***
	(0.034)	(0.036)	(0.045)
$25 \leq t \leq 30$	-0.244***	-0.301***	-0.265***
	(0.034)	(0.036)	(0.047)
$31 \leq t \leq 36$	-0.256***	-0.318***	-0.324***
	(0.036)	(0.038)	(0.059)
$37 \leq t \leq 44$	-0.184***	-0.253***	-0.283***
	(0.034)	(0.036)	(0.053)
ATE, Treatment 2:			
$1 \leq t \leq 6$	-0.157***	-0.199***	-0.136***
	(0.029)	(0.031)	(0.039)
$7 \leq t \leq 12$	-0.243***	-0.295***	-0.221***
	(0.037)	(0.038)	(0.043)
$13 \leq t \leq 18$	-0.296***	-0.355***	-0.189***
	(0.033)	(0.035)	(0.047)
$19 \leq t \leq 24$	-0.283***	-0.337***	-0.209***
	(0.035)	(0.038)	(0.045)
$25 \leq t \leq 30$	-0.235***	-0.274***	-0.187***
	(0.034)	(0.037)	(0.045)
$31 \leq t \leq 36$	-0.250***	-0.288***	-0.251***
	(0.037)	(0.039)	(0.058)
$37 \leq t \leq 44$	-0.166***	-0.219***	-0.229***
	(0.035)	(0.038)	(0.051)
Baseline waste average	0.710***	0.703***	
	(0.010)	(0.0113)	
Household-level controls	No	Yes	No
Block by period FE	Yes	Yes	Yes
Address FE	No	No	Yes
Observations	649,085	560,463	359,258
R-squared	0.339	0.336	0.001

Table presents time-varying estimates for short-run and long-run average treatment effects on per-person residual waste. Original post-intervention period covers period 1-18. 'Treatment 1' refers to monthly static-norm feedback in both Varberg and Partille. 'Treatment 2' is monthly dynamic-norm feedback in Varberg, and quarterly static-norm feedback in Partille. Within R^2 relates to remaining variation after absorbing both address and block-by-period fixed effects. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G.8: *Long-run effects of treatment*