Pre-analysis plan for project "Can children's engagement in recycling processes reduce household waste?"

This version: May 15, 2019

1. Overview of the experiment

The project is a randomized controlled trial with school children aged 10-16 in Falkenberg and Varberg municipality, Sweden. The aim of the experiment is to examine the effect of the waste-themed interventions on the waste generated by households where a child was treated. The intervention involves giving students a home assignment to measure waste amounts (treated) or weather parameters (control) over the course of one week. As part of this assignment, students also fill out paper forms with separate fields for the address where measurement occurred each day. We then couple these addresses with household-level data on collected waste amounts supplied by the municipal waste company responsible for waste management in both Falkenberg and Varberg. This allows us to identify the differential effect of the intervention on waste generation in treated versus control households.

2. Experimental design

All school visits are conducted by the same two experimenters. Each class is visited twice. On the first visit, students within a class are randomized into a treatment and a control condition. Within each condition, students are given an exercise. Treated students are asked to measure, each day over a period of one week, the amount of waste generated in their household or the household they are visiting that day. Control students face a similar task of measuring the outdoor temperature and other weather factors for a consecutive seven days.

Between one and three weeks after the conclusion of these exercises, each class is revisited by the experimenters. In this second session, treated students listen to a brief lecture on waste and the environment, participate in a subsequent group discussion, and finally play an educational game where they answer quiz questions on waste and sort cards representing different waste fractions. Control students instead listen to a lecture on geographical variation in temperature and rainfall, and participate in a similarly themed group discussion and quiz.

As part of the home assignment, each student fills in a form provided by the experimenters. This is the main data source from the intervention itself. One of the fields specifies the address where, for each day, the assignment was carried out. These addresses are then combined with household-level waste data from VIVAB, the municipal company in charge of waste management in both Varberg and Falkenberg. This allows estimation of the differential effect of treatment on waste amounts in the households where a student was treated. The form also collects information on social networks within classes, allowing us to control for social interaction as a mechanism for behavioral spillovers between treatment and control.

Schools are recruited into the study by (i) outreach with municipal managers for education, and (ii) emailing teachers directly. The schools and/or teachers that volunteer to participate are included in the study. The study is limited to the municipalities Varberg and Falkenberg because these have implemented a two-part waste tariff, where the fee paid by households depends in part on the amount thrown (in kg). The intervention is carried out with all students that are present in each class on the relevant day(s).

The study includes 33 classes averaging 20-25 student per class, summing to roughly 700-750 students. Nonresponse rates (including incomplete responses) are expected at around 20%. An additional 20% of addresses are expected to be excluded due to invalid housing type (apartment block): addresses (see below) are assigned a housing type, and all addresses coded as apartment blocks are excluded from data analysis. A small number of addresses are also likely to be excluded due to missing values in the waste data received from VIVAB. Overall, we expect to receive around 400 usable addresses.

The experimental data set has a panel structure, with several observations per participating address. The data is organized in two-week intervals running from Monday to Sunday the week after (see below), covering the period between 7 May 2018 and 17 March 2019, with the first intervention occurring on 10 September 2018 (Table 1). This implies 9 baseline (untreated) periods, 6 post-treatment periods, and 7 periods where some classes have been exposed to the intervention and some have not.

Period	Dates
1	7 May-20 May 2018
2	21 May-3 June
3	4 June-17 June
4	18 June-1 July
5	2 July-15 July
6	16 July-29 July
7	30 July-12 August
8	13 August-26 August
9	27 August-9 September
10	10 September-23 September (first class treated)
11	24 September-7 October
12	8 October-21 October
13	22 October-4 November
14	5 November-18 November
15	19 November-2 December
16	3 December-16 December (last class treated)
17	17 December-30 December
18	31 December 2018-13 January 2019
19	14 January-3 February 2019
20	4 February-17 February
21	18 February-3 March
22	4 March-17 March

Table 1. Experimental periods.

Randomization occurs along two dimensions. First within each class, students are randomized into treatment and control. This is done by manually shuffling assignment cards marked A or B. For equal numbers of students, the number of A and B cards are the same; for odd numbers, one additional A or B card is added (in an alternating pattern). Second, experimenters are randomly assigned to either group A or B (treatment or control) by means of a coin flip.

We calculate differential attrition rates, under the assumption of a prior 50-50 treatment/control split in our sample, and run a standard chi-square test of equal proportions. If this test statistic is significant, we will construct Lee bounds for the main treatment effect regression (see below). Detailed information on nonresponse rates is limited to the latter 22 classes, where we noted down the division of present students into groups A and B. However, this split is very close to 50-50 (239 vs. 241 subjects), and the same method of treatment allocation was used in the first 11 classes. Nevertheless, we also run the test only on the latter 22 classes, framing this as a robustness test.

Furthermore, sample balance will be checked across all 9 baseline periods (before 10 September 2018) with respect to averages of (i) summed residual-waste weights, (ii) summed food-waste weights, (iii) register data on number of residents at each address (see below); and (iv), with respect to the share of households in multi-family housing. For (i)-(iii), we will additionally display treatment-arm averages in a 'parallel-trend' figure across the 9 baseline periods, as is standard practice in difference-in-difference studies.

3. Data sources

1. Questionnaire data supplied by study participants (students). Contains the following information:

- (i) Name, class, school (used to construct social-network variable by third party and a class index: name/class/school strings not visible to researchers)
- (ii) Number of household members
- (iii) Age of household members
- (iv) Pet (yes/no)
- (v) List of classmates that students regularly interact with outside of school hours*
- (vi) Type of waste bin (treatment) or distance to the sea/closest lake (control)
- (vii) For each day (treatment):
  - a. The address where weighing of waste was performed*
  - b. Residual-waste weight measured by the student
  - c. Food-waste measured by the student
  - d. Whether the household was visited by someone
  - e. What the household had for dinner
- (viii) For each day (control):
  - a. The address where registration of weather was performed*
  - b. Outdoor temperature measured by the student
  - c. Whether there had been rainfall
  - d. Cloud cover (clear, cloudy, overcast)
  - e. Time of measurement

* Data used in this study.

2. Household data supplied by VIVAB. Raw data covers residual-waste and food-waste weights (in kg) at each time of collection: for details on how our variables are constructed, see the next section.

3. Register data from the Swedish Tax Authority, accessed on (and pertaining to) 26 February 2019. Used only to construct a variable for number of people per address.

4. Defining variables

Use all data on all waste bins connected with each address stated by subjects (1.vii.a and 1.viii.a), subject to restrictions given in the following section. Do not include subjects that report more than one address, but do not report any one address more than once (out of seven).

The raw data from VIVAB contains one line per bin-specific collection event. Any event with negative weights are dropped, as are lines not associated with any particular waste-bin type. Events may also generate "anomaly reports" if e.g. a bin was not placed curbside and thus could not be collected. We will consider these reports only if no strictly positive weight is reported on that data line. Weights associated with the reports that are considered may be coded as a zero or a missing value;[1] if not the anomaly report is simply dropped. See Table 2 for details on how each report is coded.

The VIVAB data sorts waste bins into three categories: food, household, and unsorted waste, where a household typically either has food and household bins, or a single unsorted-waste bin. We recode weight variables associated with household and unsorted waste as a single residual-waste variable. For each address in the data set, we then sum waste weights (in kilograms) separately for residual and food waste across the two-week periods given in Table 1.

Finally, these weights are divided the number of household members as given by the Swedish Tax Authority (data source 3), producing our two outcome variables. This last step (dividing by number of people per address) will be performed only if we are able to match at least 80% of households used in the regressions with the register data set; if not, we will use waste weights expressed in kilograms in all regressions.

5. Data restrictions

Observations fulfilling the following criteria will be considered outliers and will be dropped from the data:

- All observations of households with a mean residual- or food-waste weight above 15 kg/person
- Each observation where residual- or food-waste weight is above 60 kg/person

[1] Missing data will not be imputed.

Additionally, all observations on households that have more than 90% missing or zero observations (across all two-week periods in Table 1) for both residual and food waste are dropped.

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

Table 2. Anomaly report codes

6. Regressions and hypotheses

All regressions are difference-in-difference regressions. Because we consider the number of clusters (33 classes) too low, we do not cluster standard errors. We use randomization inference to calculate *p* values, specifically what Young (2019) terms randomization-*t* inference: we re-randomize treatment 1,000 times, calculate (regression) test statistics in each iteration, and perform inference based on the empirical distribution of the test statistics induced by the randomization distribution. Furthermore, all pairs of regressions (1-5) use critical values subjected to an adjustment for multiple hypothesis testing, the D/AP adjustment described in Sankoh et al (1997), with $K = 2$ (for regressions a and b, respectively). We will compute the correlation between residual and food waste for use in this method.

Unless otherwise noted, regressions use treatment variable T_{it}, which is always equal to zero for untreated households (see below). For treated households, $T_{it} = 1$ in all periods subsequent to the period of the first class visit by the experimenters. In the two-week period including the first visit, T_{it} is equal to the share of week days in the period occurring after the visit. Thus, for instance, if a school was visited on Thursday of the second week, $T_{it} = 0.9$. In all other periods, $T_{it} = 0$.

In some cases, a form contains a weight but no address for some date(s). In calculating modal addresses, share of days when measurement was performed at some address, etc., we ignore these fields unless only one address appears on the form in question. All measurements on that form are then assumed to have occurred at that address.

6.1 Main regression

1. Main regressions:
   a. Outcome: residual-waste weights per person (see above). Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are coded as treated as long as the treated subject reports the address at least twice in his/her form. We include only addresses that occur at least twice in some form.

   $$y_{it} = \alpha_i + \lambda_t + \beta T_{it} + \epsilon_{it}$$

   where α_i and λ_t are address and two-week period fixed effects, respectively. Note that in most cases, class, teacher, school and experimenter are all constant within addresses.

   Hypothesis: $\beta = 0$

   b. Checking mechanisms: same regression equation and treatment coding as above, but with outcome: food-waste weights per person.

   Hypothesis: $\beta = 0$

6.2 Robustness checks

2. Dropping ambiguous cases.
   a. Outcome: residual-waste weights per person (see above). Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are dropped. We include only addresses that occur at least twice in some form.

   $$y_{it} = \alpha_i + \lambda_t + \beta T_{it} + \epsilon_{it}$$

   where α_i and λ_t are address and two-week period fixed effects, respectively.

   Hypothesis: $\beta < 0$

3. Dropping period of first visit. Exactly as in regression 1a-1b, but where, for each household, we drop entirely the two-week period where the first school visit corresponding to that household occurred.

6.3 Extensions

4. Taking share of days reported into account.
   a. Outcome: residual-waste weights per person (see above). We include only addresses that occur at least twice in some form.

   $$y_{it} = \alpha_i + \lambda_t + \beta p_i T_{it} + \epsilon_{it}$$

   Here p_i is the share of days (out of seven) where measurement was performed at address i. Households that appear in more than one form are coded as treated as long as the treated subject reports the address at least twice in his/her form; p_i is then the maximum number of reported days (out of seven) among the treated subjects. Addresses that occur for some untreated subject(s) more than once in some form and for any treated subject(s) exactly once in all forms are dropped.

   Hypothesis: $\beta = 0$

   b. Checking mechanisms: same regression equation and treatment coding as above, but with outcome: food-waste weights per person.

   Hypothesis: $\beta = 0$

5. Accounting (only) for engagement with the task.
   a. Outcome: residual-waste weights per person. Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are coded as treated as long as the treated subject(s) report the address at least twice in his/her form. We include only addresses that occur at least twice in some form.

   $$y_{it} = \alpha_i + \lambda_t + \beta q_i T_{it} + \epsilon_{it}$$

   where, for each included address, q_i is the share of days (out of seven) when waste weighing occurred at some address.

   Hypothesis: $\beta < 0$

   b. Checking mechanisms: same regression equation and treatment coding as above, but with outcome: food-waste weights per person.

   Hypothesis: $\beta = 0$

6. Checking for spillovers
   a. Outcome: residual-waste weights per person (see above). Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are coded as treated as long as the treated subject reports the address at least twice in his/her form. We include only addresses that occur at least twice in some form.

$$y_{it} = \alpha_i + \lambda_t + \beta T_{it} + \gamma S_{it} + \epsilon_{it}$$

      where α_i and λ_t are address and two-week period fixed effects, respectively. S_{it} is a count variable indicating the number of classmates reported in item *(v)* in the form (i.e., classmates that students regularly interact with outside of school hours) that have $T_{it} = 1$.

      Hypothesis: $\gamma = 0$

      Exact *p* values for this hypothesis test are computed using the randomization inference method described in Athey et al. (2017). This method requires that the network matrix G be symmetric (i.e. if subject *i* interacts with subject j, subject j also interacts with i), which is not assured by our design. We therefore construct symmetric matrices by replacing element $G(i,j)$ by 1 if $G(i,j) = 0$ but $G(j,i) = 1$ (edge maximization). We then select a focal group using the greedy algorithm outlined in section 5.4.3. of Athey et al. (2017).

      We will also construct standard errors corresponding to an edge-minimizing network matrix, i.e. one where element $G(i,j)$ is replaced by 0 if $G(i,j) = 1$ but $G(j,i) = 0$; however, this analysis will be framed as a robustness test.

   b. Checking mechanisms: same regression equation, treatment coding and inference method as above, but with outcome: food-waste weights per person.

      Hypothesis: $\gamma = 0$

7. Checking for dynamics
   a. Outcome: residual-waste weights per person (see above). Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are coded as treated as long as the treated subject reports the address at least twice in his/her form. We include only addresses that occur at least twice in some form.

$$y_{it} = \alpha_i + \lambda_t + \beta_1 T_{it}^1 + \beta_2 T_{it}^2 + \epsilon_{it}$$

      where α_i and λ_t are address and two-week period fixed effects, respectively. T_{it}^1 is a treatment variable which is equal to 1 for treated households in the four weeks following the period of the first visit, with the period including the first visit coded in the same way as T_{it} in other regressions; the variable is zero otherwise. T_{it}^2 is equal to 1 for treated households starting in the fifth week after treatment, and zero otherwise.

Hypothesis: $\beta_1 = \beta_2 = 0$

This hypothesis (and others like it) will be tested using an *F* test.

b. Checking mechanisms: same regression equation and treatment coding as above, but with outcome: food-waste weights per person. Again, we adjust for multiple hypothesis testing.

Hypothesis: $\beta_1 = \beta_2 = 0$

8. Heterogeneous treatment effects with respect to baseline weights

a. Outcome: residual-waste weights per person (see above). Include the modal address given on each participant's form; in case of a tie, use both/all addresses. Households that are modal for more than one subject are dropped. Include only addresses that occur at least twice in some form.

$$y_{it} = \alpha_i + \lambda_t + \beta T_{it} + \gamma I(\bar{y}_{i0} > \bar{y}_0)T_{it} + \epsilon_{it}$$

where α_i and λ_t are address and two-week period fixed effects, respectively. $I(\bar{y}_{i0} > \bar{y}_0)$ is an indicator variable for whether the baseline residual-waste weight of household i (across periods 1-9 in Table 1) are greater than the baseline population median $\bar{y}_0$ of all household averages; households for which $I = 1$ will have treatment effect $\beta + \gamma$.

Hypothesis: $\gamma = 0$

b. Checking mechanisms: same regression equation and treatment coding as above, but with outcome: food-waste weights per person.

7. References

S. Athey, D. Eckles, and G.W. Imbens. Exact *p*-values for network interference. *Journal of the American Statistical Association*, 113(521): 230-240, 2018.

A.J. Sankoh, M.F. Huque, and S.D. Dubey. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine* 16(22): 2529-2542, 1997.

A. Young. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics* 134(2): 557-598.