Pre-Analysis Plan (ver. 2.0b)

Cluster-Randomized Controlled Trial of Sewer Connection Encouragement in Date City

Eiji Satoh
Fukushima University

Draft: May 6, 2026

Version 2.0a, April 27, 2026.
This version corrects typographical errors in Version 2.0. No substantive changes were made to the research design, randomization procedure, outcomes, hypotheses, or analysis plan.

Version 2.0b, May 6, 2026.
This version clarifies the implementation details of the constrained-randomization procedure, including the treatment of soft geographic constraints as a preferred filter within SMD-admissible allocations. No changes were made to the research questions, treatment content, outcomes, hypotheses, estimation strategy, or inference plan.

1 Working Title

Clarifying Information and Initial Behavioral Responses: A Cluster-Randomized Field Experiment on Encouraging Sewer Connections

2 Research Objective

The objective of this study is to estimate the causal effect of informational outreach on the initial behavioral responses of households that remain unconnected to the public sewer system in Date City, Fukushima Prefecture, using a cluster-randomized controlled trial.

Rather than focusing on connection completion itself, the study focuses on consultation and inquiry behavior as the key first-stage margin in the household decision process. Because sewer connection involves substantial financial costs, construction burdens, housing-related constraints, and intra-household coordination, actual connection is a medium- to long-run outcome and is therefore treated as a secondary and exploratory endpoint.

The intervention is not designed to operate through monetary incentives or strong normative pressure. Instead, it is designed to reduce informational and psychological frictions by making clear where households can seek consultation, what can be addressed through consultation, and how the first step in the process can be taken. The main question is therefore whether a low-cost information-clarification intervention increases initial contact behavior prior to later-stage actions such as estimate requests, applications, and connection completion.

More specifically, the study addresses the following questions:

1. Whether information provision that clearly identifies the consultation pathway promotes consultation and inquiry behavior.

2. Whether phased implementation, with an earlier-treatment group and a later-treatment group, generates differences in the time path of initial behavioral responses.

3. At which stage bottlenecks arise in the progression from consultation to cost estimates and application.

4. Whether treatment effects are heterogeneous across small households, households using septic tanks, and households with different prior sewer-connection outreach contact histories.

5. Whether information diffusion and spillovers through family members and neighbors can be identified in a supplementary manner using consultation logs and the follow-up survey.

3 Background and Policy Relevance

Although sewer infrastructure has expanded in Date City, a non-trivial number of households within the service area remain unconnected. This lack of connection poses challenges from the perspectives of wastewater treatment, environmental protection, and effective use of public infrastructure in line with administrative planning. At the same time, sewer connection requires households to bear in-home construction costs and related burdens, making immediate connection a difficult decision for many residents. Prior research suggests that insufficient information, psychological burdens, and unclear social norms may contribute to delayed action.

From a policy perspective, the key issue is therefore not whether connection completion can be achieved immediately, but whether early-stage actions such as consultation, inquiry, and requests for cost estimates can be encouraged.

This study evaluates the effect of a leaflet inserted into the municipal newsletter that does not simply publicize the program but instead clarifies what residents can learn by consulting and where they should make contact. The intervention is implemented in a phased manner, and the study evaluates its effect on residents' initial behavioral responses. The policy relevance of this study lies in identifying how effective such a relatively low-cost administrative intervention can be and at which stage its limitations emerge.

More broadly, this study contributes to the evidence on take-up of administrative services involving high fixed costs and procedural burdens. Rather than focusing on monetary incentives, it examines whether a low-cost information-clarification intervention delivered through an existing municipal communication channel can increase initial contact behavior before later-stage actions such as estimate requests, applications, and connection completion.

4 Overview of the Research Design

4.1 Study Area and Target Population

The primary target population consists of households in sewer service areas in Date City that are currently unconnected. For analytical purposes, households will be classified as within-area, mixed-area, or outside-area as needed, based on household-level administrative data with management IDs provided by the municipality.

4.1.1 Cluster Structure

This study uses clusters constructed from administrative *aza* units. Rather than fixing the final number of clusters ex ante, the study applies a prespecified staged clustering algorithm to the frozen roster in order to preserve geographic contiguity, reduce extremely small units, and maintain implementation-relevant local coherence. The final cluster configuration will be determined using only pre-treatment information available before assignment and will be fixed prior to the formal random assignment. A more detailed description of the cluster-construction and assignment procedure is provided in Appendix A.

4.2 Random Assignment and Phased Implementation

This study is implemented as a phased-in cluster-randomized controlled trial. The frozen roster provided in early May 2026, incorporating updates such as moves, will be used as the assignment frame, and random assignment will be conducted only once using a pre-specified random seed.

To improve comparability across groups, the final assignment will use covariate-constrained randomization under a reproducible procedure fixed prior to the intervention. The assignment procedure will seek to balance key pre-treatment cluster characteristics, including cluster size, the number of unconnected households, area composition, and wastewater-treatment composition. Pre-intervention consultation activity may be reported descriptively or used in supplementary checks, but it will not be used as a formal balance variable for the final assignment. The procedure will impose hard constraints on overall early-wave/later-wave balance and district-level balance, while treating approximate balance within larger oaza and symmetric allocation within very small oaza as pre-specified soft geographic constraints. Within the set of SMD-admissible allocations, assignments with no soft-constraint violations will be given priority when such assignments exist; if none exist, the soft constraints will be relaxed according to the pre-specified fallback rule. Among the resulting selection set, one allocation will be selected using a pre-specified random seed. Additional details on cluster construction and assignment are provided in Appendix A.

The study period and treatment timing are as follows:

Period	Status
April–June 2026	No clusters treated (baseline period)
July–September 2026	First-wave clusters treated
October–December 2026	Second-wave clusters treated

Baseline consultation activity will be defined using the pre-intervention observation window from April to June 2026. Because this window is short and consultation activity is conceptually close to the primary outcome, consultation counts and rates from this period will not be used as formal balance variables for the final assignment. They may instead be reported descriptively or used in supplementary checks, subject to data quality.

Any testing or rehearsal of the randomization program conducted prior to May is intended only to verify data consistency and the reproducibility of the assignment procedure. Such exercises do not constitute the formal random assignment used in the study.

4.3 Intervention

The intervention consists of distributing a sewer connection encouragement leaflet inserted into the municipal newsletter. The finalized leaflet is based on existing materials but revised to clarify the consultation pathway, explicitly state that consultation alone is welcome, organize what residents can confirm first regarding costs and construction, improve readability for older residents, and make consultation contacts and contact methods more specific.

The intervention is intended to reduce informational frictions by clarifying the consultation pathway and lowering the psychological cost of taking the first step. Relative to business-as-usual administrative practice, the treatment adds a low-cost information-clarification component through an existing municipal communication channel, while ordinary consultation, application, and related administrative procedures remain unchanged and available to all households.

4.4 Hypothesized Behavioral Mechanism

The hypothesized behavioral pathway is as follows:

Information provision
→ Understanding of the consultation pathway
→ Reduced anxiety and recognition that consultation can clarify the next step
→ Consultation or inquiry
→ Cost estimate request or contractor referral
→ Application for connection
→ Connection completion

5 Data

5.1 Baseline Administrative Data (Household Master File)

The study uses household-level administrative data with management IDs from which personally identifying information has been removed.

The main variables are expected to include:

- Management ID

- Oaza ID
- Aza ID
- Area-use classification
- Current wastewater treatment classification
- Household characteristics (e.g., household size)
- Prior outreach history

5.2 Behavioral History Data

The municipality will provide household-level records of the dates on which the following events occurred:

- Consultations or inquiries
- Cost estimate request
- Application
- Completion

The observation period begins in April 2026, and earlier records will be incorporated if available. The initial planned end date for data provision is March 2027. If possible, the period will be extended through December 2027 in order to capture medium- to long-term outcomes such as applications and connection completion.

As a general rule, the primary outcome will be measured using these behavioral history data.

5.3 Consultation Log Data

To understand consultation content and mechanisms, the study will use consultation logs in which each row corresponds to one consultation. The main variables include:

- Management ID
- Consultation date
- Mode of contact
- Staff ID responding to the consultation
- Type of respondent
- Trigger for consultation
- Information heard from family or neighbors
- Whether the leaflet was seen

- Concern before consultation
- Consultation content
- Consultation level
- Whether a contractor was contacted
- Next step
- Concern after consultation
- Research exclusion flag

Consultation logs will be used both to support measurement of the primary outcome and to analyze mechanisms, consultation content, and other supplementary outcomes.

5.4 Follow-Up Survey

To understand mechanisms among non-consulting households that do not appear in the consultation logs, a brief follow-up survey is planned for December 2026, after the main short-run evaluation period. The follow-up survey is not part of the primary analysis; rather, it is treated as a supplementary module for mechanism analysis.

The target population consists of households with no consultation record by the end of November 2026. Sampling will be based on stratified random sampling, including balanced representation from the first-wave and second-wave groups and, if necessary, oversampling mixed-area aza units. The sampling procedure and random seed will be fixed in advance.

The main items to be measured in the follow-up survey include:

- Leaflet recognition
- Understanding of where to seek consultation
- Recognition of whether the household is in the target area
- Main reasons for not consulting
- Decision-making constraints related to cost, procedures, or uncertainty
- Intention to consult in the future

The detailed questionnaire will be finalized in an appendix or supplementary version. The planned mailing scale is approximately 600 households, subject to administrative feasibility, budget constraints, and the expected response rate.

The follow-up survey will not be used to redefine the main analysis sample or the primary outcome, and it will not be treated as the main source for causal identification. Because survey response is voluntary and nonresponse may be non-random, findings from the follow-up survey will be interpreted as supplementary evidence for mechanism analysis rather than as the primary basis for causal inference.

6 Definition of Outcome Measures

6.1 Primary Outcome

In the primary analysis, outcomes are defined at the household-by-month level.

- `Any_Consult_it`: Equals 1 if household i has at least one sewer-related consultation or inquiry recorded in month t, and 0 otherwise.

`Any_Consult_it` includes sewer-related consultations or inquiries that can be linked to the household-level administrative record according to pre-specified rules. Contacts that cannot be linked at the household level, but can be tracked only at the cluster level, will be used in supplementary descriptive analyses rather than in the main household-level outcome definition.

6.2 Secondary Outcomes

- `Estimate_Request_it`: Whether a cost estimate was requested during period t (0/1)
- `Application_it`: Whether a connection application was submitted during period t (0/1)
- `Contractor_Contact_it`: Whether a contractor was contacted (0/1)
- `Consultation_Level_it`: Consultation level, as assessed by staff
- `Concern_After_Consultation_it`: Concern status after consultation based on the consultation log
- `Pathway_Clarity_it`: Clarity of the consultation pathway based on the consultation log
- `Number_of_Contacts_it`: Number of consultations

Among these, `Estimate_Request_it`, `Application_it`, and `Contractor_Contact_it` are treated as the main secondary outcomes. The others are treated as supplementary or mechanism-related outcomes.

6.3 Exploratory Outcomes

- `Connection_Completed_it`: Whether sewer connection construction was completed
- Supplementary composite indices of concern and pathway understanding based on consultation logs
- Changes in connection rates at the aza level or cluster level
- Spillover indicators through family and neighbors, based on the consultation-log item "information heard from family or neighbors" and information-source items in the follow-up survey
- Leaflet recognition, target-area recognition, understanding of where to seek consultation, anxiety, and consultation intention measured in the follow-up survey

7 Estimation Strategy

7.1 Main Estimation

The primary estimand in this study is the intention-to-treat (ITT) effect on household-level binary outcomes, such as whether a consultation occurred in a given month. The main estimator is a linear probability model (LPM) with cluster fixed effects and month fixed effects.

The main specification is the minimal model including only cluster fixed effects and month fixed effects. A covariate-adjusted specification will also be estimated as a supplementary specification intended to improve precision.

Let Y_{ict} denote the outcome for household i in cluster c and month t, and let $Treat_{ct}$ indicate whether cluster c has already received the intervention by month t. The key parameter of interest is the coefficient on $Treat_{ct}$, which represents the average effect of treatment assignment on the probability of consultation and inquiry behavior.

The covariate-adjusted specification includes only a small set of attributes that can be measured relatively stably prior to the intervention. Specifically, these include area classification, current wastewater treatment classification, household size, and prior sewer-connection outreach contact status (e.g., direct in-person contact, mail-only contact, unknown). Household-level consultation history from April to June 2026 is conceptually close to the primary outcome, but because it may depend on the short observation window and data quality, it will not be included as a household-level covariate in the main specification. Instead, it may be used in supplementary specifications or descriptive statistics. Both models, with and without the pre-specified covariate vector, will be reported in the main results tables.

The estimating equations are:

Minimal specification (main specification)

$$Y_{ict} = \alpha + \beta Treat_{ct} + \gamma_c + \delta_t + \varepsilon_{ict} \tag{1}$$

Covariate-adjusted specification

$$Y_{ict} = \alpha + \beta Treat_{ct} + \gamma_c + \delta_t + X_{ic}'\theta + \varepsilon_{ict} \tag{2}$$

Here, γ_c denotes cluster fixed effects, δ_t denotes month fixed effects, and X_{ic} denotes the vector of pre-specified covariates.

The LPM is used because coefficients can be interpreted directly as probability differences and because inference remains transparent in fixed-effects settings. Since treatment is assigned at the cluster level, and since transparency and interpretability are particularly important with a small number of clusters, the main specification uses cluster fixed effects.

All inference will cluster at the unit of randomization, that is, the constructed cluster. Given the limited number of clusters, the main inference will rely on CR2 (bias-reduced linearization) variance estimators with Satterthwaite-adjusted degrees of freedom. As robustness checks, the study will also report p-values based on the wild cluster bootstrap and, where feasible, randomization-inference-based p-values that respect the realized constrained-randomization procedure.

Because events may be relatively rare, supplementary analyses will estimate Poisson models or, if necessary, negative binomial models using cluster-by-month consultation counts. In these models, the number of target households will be included as an offset.

Nonlinear models for binary outcomes, such as logit models, will not be treated as the main specification because of the complexity of fixed-effect handling and inference. They may be reported as sensitivity analyses if needed.

To address multiple testing, the primary outcome `Any_Consult_it` will be evaluated separately without adjustment. For the pre-specified key secondary outcomes—`Estimate_Request_it`, `Application_it`, and `Contractor_Contact_it`—the study will control the false discovery rate using the Benjamini-Hochberg procedure, applied to the p-values from the main CR2-based inference. The remaining secondary outcomes, exploratory outcomes, and subgroup analyses will be treated as supplementary and interpreted with appropriate caution rather than as the basis for the main confirmatory claims. As a robustness check, Holm-adjusted p-values for the key secondary outcomes will also be reported in an appendix.

The primary analysis targets the assignment-based ITT effect. Accordingly, even if there are minor deviations in timing or imperfect implementation, the main analysis will remain based on the ex ante randomized assignment.

Identification comes primarily from the contrast between first-wave and second-wave clusters during the period from July to September 2026, when first-wave clusters have already received the leaflet and second-wave clusters have not yet received it, while controlling for cluster fixed effects and month fixed effects. The post-October period, when both waves have received the intervention, is used primarily to trace the subsequent time path of outcomes rather than as the main source of identification.

As a supplementary descriptive analysis, the study will also examine event-time patterns around leaflet distribution to assess the timing, persistence, and possible anticipation of treatment effects.

7.2 Subgroup Analysis

The study will examine heterogeneity in treatment effects using pre-specified subgroup analyses. These analyses are intended to complement the main analysis, and some are exploratory in nature.

The main heterogeneity dimensions of interest are:

- Mixed aza units versus fully covered aza units
- Small households
- Prior sewer-connection outreach contact status (direct in-person contact / mail-only / unknown)

Exploratory heterogeneity analyses will consider:

- Households using septic tanks
- Clusters with low versus high baseline connection rates

- Clusters with shorter versus longer time since sewer service became available
- Clusters with lower versus higher pre-intervention consultation activity

Subgroup analyses will be implemented by augmenting the main estimating equation with subgroup indicators and their interactions with the treatment indicator. For example, when examining heterogeneity by a subgroup attribute G, the model will include an interaction term $Treat_{ct} \times G_{ic}$, and the coefficient on the interaction will be interpreted as the differential treatment effect across subgroups.

Inference in subgroup analyses will also account for within-cluster correlation and will use methods appropriate for a small number of clusters.

Because subgroup analyses may face substantial uncertainty due to multiple testing and limited sample size, interpretation will emphasize point estimates and confidence intervals rather than statistical significance alone.

7.3 Missing Data and Data Processing

This study uses administrative data with management IDs and consultation logs provided by the municipality. The following rules for data processing and missing data will be specified in advance.

(1) Behavioral history data

If no record of consultation, estimate request, application, or completion exists during the relevant period, the corresponding behavior will in principle be coded as not having occurred, under the maintained assumption that the administrative logs were collected according to the intended procedure. However, because omissions and recording errors are possible, the researcher will review monthly submissions for sudden changes in counts, implausible missingness, and increases in consultations with unknown management IDs.

(2) Partial missingness in consultation logs

Some fields in the consultation log may be left unrecorded. In such cases, the municipality will, where possible, distinguish between "not confirmed" and "no response" rather than leaving blank cells. In analysis, these categories will either be treated as distinct categories or handled as missing as appropriate.

(3) Consultations with unknown management IDs

Consultations for which the management ID cannot be identified will still be recorded in the consultation log where possible. In such cases, aza ID, administrative district ID, consultation date, and other available information will be used for auxiliary tracking. If the consultation cannot be linked rigorously to household-level data, it will not be used in the main analysis. However, it may be used as a reference in cluster-by-month count aggregation or descriptive statistics.

(4) Research exclusions (opt-out)

If a resident indicates that they do not wish their data to be used for research, the municipality may still retain the record for ordinary administrative purposes, but the consultation log will include a research exclusion flag. The researcher will exclude observations with this flag from the analysis sample. The number of exclusions and any group differences in exclusions will be reported descriptively.

(5) Updates to cluster definitions and area classifications

Because moves and administrative updates may affect aza, administrative district, and area classification information, the main analysis will rely on the cluster definitions and area classifications in the frozen roster provided in early May 2026. If additional corrections become available later, they will be processed according to pre-specified rules so as not to affect the main analysis in an ad hoc manner, and alternative tabulations may be reported as robustness checks if needed.

(6) Missing data in the follow-up survey

The follow-up survey is treated as supplementary and is not used in the primary estimation. Therefore, nonresponse in the follow-up survey does not affect the main analysis sample. Analyses using the follow-up survey will report the distribution of respondent characteristics and response rates, and, where appropriate, compare respondents and nonrespondents. Findings from the follow-up survey will be interpreted as supplementary evidence for mechanism analysis rather than as the primary basis for causal inference.

(7) Recording implementation fidelity and noncompliance

To assess whether the intervention was implemented as planned, the study will record, where feasible, the number of leaflets printed, printing costs, the number of target recipients, the number actually distributed, remaining copies, the date or month of distribution, the version of the leaflet used, whether the distribution timing changed, whether monthly consultation logs were submitted, and whether monthly behavioral history data were submitted. These are not primary outcomes, but they will be used as supplementary information for assessing implementation fidelity and interpreting the results. If delayed distribution, non-distribution, changes in leaflet version, or failures to submit monthly logs occur, these will be described as implementation deviations and may be considered in supplementary sensitivity analyses. However, the main analysis will remain based on the ex ante randomized assignment and the ITT framework.

(8) Recording implementation costs

To evaluate the policy feasibility of the intervention, the study will record, where feasible, the number of leaflets printed and printing costs, whether there were any additional distribution expenses, approximate administrative staff time required, approximate per-consultation time required to enter consultation logs, and any additional costs and staff time associated with the follow-up survey.

This is not intended to be a full cost-benefit analysis, but rather a descriptive record of implementation costs to assess practical scalability and feasibility. Where feasible, the study will also report descriptive measures such as total additional administrative workload, total additional implementation cost, and approximate cost per additional consultation. Where feasible, the municipality will also maintain a separate simplified time-use record for consultation handling during limited pre-specified observation windows. This supplementary record will be maintained separately from the main consultation log and will include approximate consultation start time, end time, duration, and handling complexity. It will be used only as a descriptive measure of additional administrative workload and implementation cost, not as an outcome measure, a staff performance record, or an input into the primary causal analysis.

(9) Supplementary use of GIS-based proximity information

GIS-based indicators of adjacency or distance across clusters will be used primarily for supplementary evaluation of spillovers and information diffusion. These variables will not be included as household-level covariates in the main estimation. Instead, they will be treated as descriptive or supplementary analytical measures. For example, they may be used as adjacency indicators, measures of proximity to already treated clusters, or indicators for assessing whether nearby clusters are disproportionately assigned to the same treatment status.

(10) Supplementary evaluation of spillovers

Information diffusion and spillovers through family members and neighbors will be evaluated primarily in a descriptive and supplementary manner using the consultation-log item "information heard from family or neighbors," information-source items in the follow-up survey, and GIS-based proximity indicators. Because rigorous identification at the household level may be difficult, these analyses will not be treated as central to the main causal estimates.

(11) Data processing and reproducibility

The procedures for data cleaning, variable construction, and creation of the analytical sample will be coded and archived in reproducible form by the researcher. In addition, the researcher will manage, to the extent feasible, the raw data receipt files, anonymized analysis datasets, data-cleaning code, randomization code, analysis code, figure/table production code, and codebook as distinct components.

8 Sample Size and Statistical Power

The number of clusters, the number of target households, and the number of unconnected households will be finalized based on the frozen roster provided in early May 2026. A formal power calculation based on the frozen roster will be added once the frozen roster is received and before treatment begins. The calculation will use the finalized number of clusters, cluster sizes, and plausible assumptions about baseline consultation rates and intracluster correlation for the primary outcome. If the resulting minimum detectable effect is larger than initially anticipated, interpretation will place greater weight on effect magnitudes, confidence intervals, and policy relevance than on statistical significance alone.

In this setting, even modest percentage-point increases in consultation behavior may be policy-relevant if they can be achieved at low implementation cost through an existing municipal communication channel.

9 Pre-Registration

This study will be registered in the AEA RCT Registry prior to the start of the intervention. Related documents may also be archived on OSF or similar platforms as needed. The registration will include the research objective, intervention content, primary and secondary outcomes, unit of randomization, and analysis plan. The follow-up survey will also be included in the registration in terms of whether it will be conducted, its target population, the sampling approach, and the key measurement items. Detailed questionnaire wording may be added in a supplement if needed.

10 Research Ethics

- The researcher will not receive personally identifying information such as names or addresses.
- The study will use municipality-provided data with management IDs and consultation logs.
- The correspondence table linking management IDs to personally identifying information will be retained by the municipality, and the researcher will not have access to it.
- If a respondent indicates, through the consultation log or the follow-up survey, that they do not wish their data to be used for research, the corresponding observation will be excluded from the analysis based on a research exclusion flag.
- The researcher will not attempt to re-identify any individual or household.
- The study will be conducted following approval by the relevant university research ethics review process.
- In all publications and reports, results will be presented only in a form that does not allow identification of any individual household or person.
- Consultation, application, and other administrative procedures will proceed as usual regardless of whether a household agrees to research use of its data.

11 Limitations and Points to Note

- Final cluster definitions and treatment assignment will be determined based on the frozen roster provided in early May 2026.

- The planned treatment timing is July 2026 for the first-wave group and October 2026 for the second-wave group.
- The leaflet used for the intervention has already been finalized and will be attached as an appendix.
- The follow-up survey is intended as a supplementary mechanism module, and its detailed wording and field implementation procedures will be finalized separately and documented in a supplementary version if needed.
- The end date of behavioral history data provision depends on coordination with the municipality, and an extension may be requested if necessary to capture applications and completion outcomes.
- Records on implementation costs, including any simplified consultation time-use records, will be based on limited observation windows and simplified documentation to the extent feasible.

12 Planned Main Tables and Figures

To enhance transparency in reporting results, the study plans to produce at least the following tables and figures.

Tables

1. Descriptive statistics for the analysis sample
2. Pre-randomization balance table
3. Main estimation results for the primary outcome
4. Main estimation results for key secondary outcomes
5. Results of heterogeneity analyses
6. Descriptive table on implementation fidelity and implementation costs
7. Descriptive table on final cluster definitions and assignment constraints

Figures

1. Monthly trends in consultation rates or consultation counts
2. Event-time style supplementary figures, if needed
3. Map of final clusters and phased assignment timing, subject to disclosure constraints

The exact structure of the tables and figures will be finalized after the analytical dataset is completed, but the reporting will clearly distinguish primary outcomes, secondary outcomes, and implementation measures.

13 Appendices

A. Cluster Construction and Assignment

B. Distribution Leaflet

C. Data Request Letter

D. Data Provision Request Specifications

E. Front-Desk Consultation Log Items and Operational Notes

F. Research Timeline

G. Draft Follow-Up Survey Plan (to be attached once finalized)

H. Implementation Cost and Supplementary Time-Use Record Sheet (to be attached if needed)

A Cluster Construction and Assignment

Cluster Construction and Assignment

For the Sewer Connection Encouragement RCT only[1]

The geographic building block for cluster construction is the administrative *aza*. Clusters are constructed using a prespecified staged algorithm designed to preserve geographic contiguity and implementation coherence while avoiding extremely small clusters. Adjacency is coded ex ante using administrative *aza* identifiers and map-based review. Adjacency categories distinguish non-adjacent pairs, weak-contact pairs, adjacent pairs separated by strong barriers, and adjacent pairs without strong barriers; the clustering algorithm prioritizes the last category. Non-contiguous mergers are not permitted.

Cluster construction proceeds in stages. First, we merge extremely small *aza*-level units into natural contiguous neighbors where feasible. Second, we attach remaining small residual units to geographically natural anchor clusters. Third, we consolidate medium-sized units only when doing so yields geographically coherent clusters of reasonable size. Units for which no natural contiguous merger exists are retained as exceptional stand-alone clusters. We do not fix the final number of clusters ex ante. Instead, we apply the same prespecified rules to the frozen roster and adopt the cluster configuration that arises naturally from those rules.

After final clusters are defined, we assign clusters to the early-wave and later-wave groups using covariate-constrained randomization. We generate a large number of candidate assignments and retain only those satisfying the hard constraints, namely overall balance in the number of clusters assigned to each wave and balance at the district level. For each hard-constraint-satisfying assignment, we compute balance metrics based on pre-specified cluster-level covariates, including the number of unconnected households, cluster size, area composition, and wastewater-treatment composition. For each balance variable, we compute the standardized mean difference between the early-wave and later-wave groups, and we use the maximum absolute standardized mean difference and the sum of squared standardized mean differences as the main balance metrics.

We define SMD-admissible assignments as hard-constraint-satisfying assignments whose maximum absolute standardized mean difference is no greater than the pre-specified threshold. Approximate balance within larger *oaza* and symmetric allocation within very small *oaza* are treated as soft geographic constraints. These soft constraints are not hard constraints, but they are used as a pre-specified preferred filter: if at least one SMD-admissible assignment has no soft-constraint violations, the final selection pool is restricted to such assignments. If no SMD-admissible assignment satisfies all soft constraints, the soft constraints are relaxed according to the pre-specified fallback rule while all hard constraints are maintained. If multiple assignments remain in the final selection pool, we form a selection set from the assignments with the lowest values of the pre-specified overall imbalance index and select one assignment using a pre-specified random seed.

All adjacency rules, cluster-construction rules, exception handling, allocation constraints, bal-

[1]This appendix is a research-oriented English version prepared for the PAP. This appendix is prepared for the sewer connection encouragement RCT only. Although some administrative recording procedures may in practice be shared with the combined septic tank conversion promotion RCT, the present appendix specifies the cluster construction and assignment procedure for the sewer connection encouragement RCT and should be interpreted accordingly.

ance metrics, soft-constraint fallback rules, and seed management are finalized before receipt of the frozen roster and are not revised afterward. The formal assignment is conducted once, using the frozen roster only. A more detailed implementation protocol was finalized before receipt of the frozen roster and is retained by the researcher as the internal assignment record.

B English Rendering of the Distribution Leaflet

Connecting to the Public Sewer System

Where to Ask for Advice

Thank you for your continued understanding and cooperation regarding Date City's sewerage services.

To improve the living environment and protect water quality, households that may be eligible for connection to the public sewer system are encouraged to first contact the **Waterworks and Sewerage Division of Date City** for advice. Consultation by phone is also available.

- It is perfectly fine to contact the City for advice even if you have not yet decided whether to proceed with construction.
- After confirming your situation, the City can provide a list of designated drainage equipment contractors.
- Approximate costs and construction timing can then be confirmed with a designated contractor.
- If you prefer, you may also contact a designated contractor directly.

Contact

- **Phone:** 024-573-5059
- **Office hours:** Weekdays, 9:00 a.m. to 5:00 p.m. (excluding city holidays)
- **Address:**
  Waterworks and Sewerage Division, Construction Department, Date City
  180 Funahashi, Hobara-machi, Date City, Fukushima 960-0692
- In-person consultation is also available.

Example of What to Say on the Phone

> "I would like to ask about connecting to the sewer system.
> I would like to know the approximate cost and the general construction process."

When you contact the City, it is helpful if you can tell us:

- your district (e.g., aza / neighborhood area), and
- if known, the current status of your toilet and wastewater drainage.

Family members may also contact the City on your behalf.

Important Construction work is required on the household side of the sewer connection. Please contact the City first if you would like to confirm the approximate cost and the general process.

Note: This flyer is distributed to neighborhood associations that include areas where sewer service is available. It may also reach locations outside the service area, places where service has not yet started, or households that are already connected.

Frequently Asked Questions

Q1. If I contact the City for advice, do I have to proceed with the construction?
A. No. Asking for advice does not mean that you are applying for connection. It is perfectly fine to begin by simply confirming your situation.

Q2. What should I tell the City when I ask for advice?
A. It is sufficient to tell us where you live and, to the extent you know, the current status of your toilet and wastewater drainage.

Q3. About how much will it cost to connect?
A. Construction costs vary depending on each household's situation. A designated contractor will provide an estimate after confirming the conditions at your property. Depending on the circumstances, interest subsidies or similar support may be available. Eligibility can be confirmed when you consult the City.

Q4. If I obtain an estimate, do I have to proceed with the construction?
A. No. Even if you obtain an estimate, you are not required to proceed.

Q5. I do not really understand what kind of construction would be involved.
A. The City will first listen to your situation and, if appropriate, provide a list of designated drainage equipment contractors.

Q6. Can older residents complete the procedure?
A. Yes. Family members may consult or complete procedures on their behalf.

Summary

1. Contact the City for advice
2. Receive a list of designated contractors
3. Obtain an estimate from a designated contractor

Consultation does not mean application. It is perfectly fine to begin by confirming your situation first.

Research Notice

To make the City's guidance easier to understand, Date City is cooperating with the laboratory of Eiji Satoh at Fukushima University in compiling and analyzing summary information on consultations and related responses.

When results are shared outside the City, no names, addresses, or other personally identifying information will be disclosed. You may be asked to respond to brief confirmation questions or a short survey, but participation is voluntary.

Whether or not you cooperate in the research, consultation, application, and related administrative procedures will proceed as usual.

Research Use of Information

If you do not wish your information to be used for research, please contact the Waterworks and Sewerage Division.

Contact for research-use opt-out:
Waterworks and Sewerage Division, Construction Department, Date City
Phone: 024-573-5059

C Request for Data Provision

Request for Data Provision
For the Sewer Connection Encouragement RCT only[2]
Prepared as an appendix to the PAP
To: Date City Waterworks and Sewerage Division

C1. Purpose of this request

In order to conduct a randomized controlled trial (RCT) evaluating the effects of information provision for sewer connection encouragement in Date City, it is necessary to understand household characteristics as well as behaviors such as consultations, inquiries, estimate requests, applications, and connection completion, and to conduct scientific analyses that can inform policy improvement.

This document requests the provision of management-number-based household-level data and consultation logs recorded during the implementation period of the RCT. The researcher will not receive any personally identifiable information such as names or addresses.

C2. Data Requested

For each household in Date City, we kindly request the provision of the items below. All data should be provided based on management numbers, with personal information removed.

A. Household Master Data (Core Items)

Item	Description
Management number	Management number assigned by the City (identifier that does not contain personal information)
Oaza/Koaza ID	Area code or name at the small-area level
Administrative district ID	Area code or name at the administrative district level
Household size	Number of household members
Land-use category	Code for land-use category
Wastewater treatment category	Code for wastewater treatment category
Public sewer service district	Code for public sewer service district

If an anonymized household ID is already in use, it may also be included provided that it can be linked consistently to the management number.

[2] This appendix is a research-oriented English version prepared for the PAP. The substantive data items and administrative requests are based on the Japanese data request letter submitted to Date City. This document is prepared as a supporting appendix to the PAP for the sewer connection encouragement RCT. Version numbers for the sewer PAP are tracked independently from those for the combined septic tank conversion promotion RCT.

B. Data Required for the Sewer Connection Encouragement RCT

Item	Description
Past sewer connection encouragement activities	Past encouragement activities, to the extent identifiable (e.g., year, target households, whether direct visits were conducted, whether mail-only outreach was used, and other available outreach classifications)
Household-level behavioral history	Dates of consultations/inquiries to the City, estimate requests, applications, and completion of sewer connection
Consultation log	Consultation-level records (see Section C below)

C. Items Requested in the Consultation Log (One Row per Consultation)

For the consultation log, we kindly request that one row be recorded for each consultation. Because multiple topics may arise in a single consultation, multiple selections may be allowed where appropriate. If possible, it would be helpful to use the same format consistently across related wastewater-policy consultations handled by the Division. The main requested items are as follows:

- Consultation ID
- Management number (if known)
- Date of consultation
- Reception channel
- Staff ID handling the consultation (anonymous code)
- Respondent category
- Oaza/Koaza ID
- Administrative district ID
- Sewer service area classification
- Current wastewater treatment status
- Trigger of the consultation
- If prompted by family or neighbors, what information they heard
- Whether the flyer was seen
- The main concern before consultation
- Consultation topics

- Consultation level (staff judgment)
- Whether the household had contacted a contractor
- Next step (staff judgment)
- Level of concern after consultation
- Clarity of the action pathway
- Research exclusion (opt-out)

Detailed consultation-log items and input procedures are specified in the attached *Data Provision Specification* and *Consultation Manual*.

C3. Data Format

We would appreciate it if the data could be provided in the following format.

(1) Household Master Data (Household Level)

One row per household; one column per variable.
Primary key: management number.

(2) Consultation Log (Consultation Level)

One row per consultation; one column per variable.
Linked to the household master by management number.

If the management number is unknown, the field may be left blank; if possible, Oaza/Koaza ID and administrative district ID may be used as supplementary information.

Please provide the data in Excel format. As a rule, it would be helpful if the workbook were organized into separate sheets for *Household Master* and *Consultation Log*. If this is difficult in practice, separate files are also acceptable. If possible, please also attach item descriptions (code list and/or data dictionary).

For updated data during the RCT implementation period, we kindly request monthly submissions on a month-end basis beginning in July 2026.

C4. Requested Data Coverage Period

For the analyses of the sewer connection encouragement RCT, we would like to request data covering the following periods.

(1) Baseline Data

- Most recent household master data
- Household attributes (e.g., land-use category, wastewater treatment category, household size)
- Area classification
- Management number

(2) Past Behavioral History

If possible, for a pre-intervention reference period:

- Sewer-related consultations and inquiries
- Estimate requests
- Applications
- Completion records
- Past encouragement activities, to the extent identifiable

(3) Behavioral History and Consultation Logs During the RCT Implementation Period

As a baseline request, we ask that data be provided for the period from April 2026 to March 2027.

- Consultations / inquiries
- Estimate requests
- Applications
- Completion records
- Consultation logs (one row per consultation)

(4) Request for Possible Extension

To capture medium- and long-term outcomes such as applications and connection completion, we may ask, if feasible, for the data coverage period to be extended through December 2027.

C5. Protection of Personal Information and Research Ethics

All data will be provided only after personal information has been removed. The researcher will not receive names, addresses, or other personal identifiers.

The study will use municipality-provided data with management numbers and consultation logs. The correspondence table linking management numbers to personally identifying information will be retained by the City, and the researcher will not have access to it.

If a resident indicates, through the consultation log or a follow-up survey if implemented, that they do not wish their data to be used for research, ordinary administrative records will still be retained as needed for routine administrative purposes, but a research-exclusion flag will be added so that the researcher can exclude the corresponding record from analysis.

The researcher will not attempt to re-identify any individual or household. This study will be conducted after approval by the university research ethics review process. Data will be managed strictly and used only for research purposes. Only aggregated results will be reported in papers and reports; no individual household or person will be identifiable.

Consultation, application, and other administrative procedures will proceed as usual regardless of whether a household agrees to research use of its data.

C6. Follow-Up Survey

Because consultation logs cannot capture the situation of households that do not consult, we may separately request the implementation of a brief follow-up survey targeting households with no consultation record during the relevant evaluation period.

If the follow-up survey is implemented, the anticipated division of responsibilities is as follows:

- The researcher will use the management-number-based data to select the sample through stratified random sampling.
- The researcher will provide the City with a list of management numbers corresponding to the households to be mailed.
- The City will be responsible for attaching addresses, mailing the survey, and receiving the responses.

Details will be formally requested after further consultation. The consultation logs and follow-up survey will be used solely for analyses aimed at improving policy implementation, and they will not affect residents' procedures or the availability of municipal services.

In addition, to document the approximate administrative workload associated with the intervention, the researcher may request, to the extent feasible, a simple time-use record for telephone or in-person consultations during a limited observation window. This record is intended only as a supplementary measure of the administrative burden associated with the intervention and not as a detailed work log, attendance record, or staff performance record. Such recording may be maintained separately from the main consultation log, including in a separate worksheet within the same Excel workbook.

C7. Burden on the City and Support from the Researcher

The researcher will respond flexibly to questions about data formats and variable definitions. The study will be designed so as to minimize the administrative burden on the City. The researcher will prepare draft formats for the consultation log, data-entry rules, code tables, and, if needed, draft sample-selection lists for the follow-up survey. The researcher can also provide explanations at internal municipal meetings if needed.

C8. Contact

Eiji Satoh, Fukushima University
024-548-8452
esatoh@econ.fukushima-u.ac.jp

Sincerely,

D Data Provision Request Specifications

Data Provision Request Specifications
For the Sewer Connection Encouragement RCT only[3]
Prepared as an appendix to the PAP

D1. Purpose of This Document

This document specifies the data items and delivery format requested from Date City for the Sewer Connection Encouragement RCT. It is intended to be read together with the main PAP and the related appendices, and it corresponds to the data structure, behavioral records, consultation log, and analysis plan described in the PAP.

This specification applies to:

- the first batch of data previously created by the City,
- the updated dataset to be compiled and provided in early May 2026, and
- the monthly submission data during the implementation period of the RCT.

D2. Basic Data Structure

The data used in this research consist primarily of the following four components:

(A) **Household-level baseline data [Required]**
A household master file organized with one row per household and one variable per column. The primary key is the administrative management number.

(B) **Behavioral history (household-level)**
Dates of events such as consultation/inquiry, estimate request, application, and completion recorded at the household level.

(C) **Supplemental data [If available]**
Past outreach history, past consultation records, and related information that may be useful for analysis.

(D) **Consultation log (consultation-level) [Required]**
A consultation record organized with one row per consultation and one variable per column. It is intended to be linked to the household master by the administrative management number.

[3] This appendix is a research-oriented English version prepared for the PAP. The substantive data items and administrative requests are based on the Japanese data provision request specifications letter submitted to Date City. This appendix is prepared for the sewer connection encouragement RCT only. Although some administrative recording procedures may in practice be shared with the combined septic tank conversion promotion RCT, the present document specifies the data required for the sewer RCT and should be interpreted accordingly.

D3. Household-Level Baseline Data (A)

Basic Structure

- One row per household
- One variable per column
- Primary key: administrative management number

Main Items

Item	Description	Format (example)
Administrative management number	Household identifier that does not contain personal information	Numeric or text
Oaza/aza ID	Area code or name at the aza level	Numeric or text
Administrative district ID	Area code or name at the administrative district level	Numeric or text
Household size	Number of household members	Numeric
Land-use category	Land-use category code	Code
Wastewater treatment category	Wastewater treatment category code	Code
Public sewer treatment subdistrict	Public sewer treatment subdistrict code	Code
Sewer service area classification	Inside service area, outside service area, mixed, etc.	Code
Current treatment status	Combined septic tank, single septic tank, night-soil collection, etc.	Code

If an anonymized household ID is already used internally, it may also be included as long as it can be mapped consistently to the administrative management number.

D4. Behavioral History (B)

Basic Structure

- One row per household
- Event dates for consultation/inquiry, estimate request, application, completion, and related events are managed in columns
- If the same household has multiple occurrences, multiple columns may be used as needed (e.g., Consultation Date 1, Consultation Date 2)

Coverage Period

- Historical records: as available, from April 2023 to March 2026
- RCT period: April 2026 to March 2027 as the baseline requested period
- To capture medium- to long-term outcomes such as application and completion, an extension through December 2027 may be requested if feasible

Main Items

Item	Description	Format (example)
Sewer Consultation Date 1	First consultation / inquiry date	Date
Sewer Consultation Date 2	Second consultation / inquiry date	Date
Sewer Estimate Request Date 1	First estimate request date	Date
Sewer Application Date 1	First application date	Date
Sewer Completion Date 1	First completion date	Date

Column names may be flexibly set to match the actual administrative files. The researcher will standardize them as needed.

D5. Supplemental Data (C) [If Available]

If available, the following supplemental data would be useful for analysis:

- Past sewer connection encouragement records
- Past consultation history related to sewer connection
- Additional information on area or household characteristics
- Any administrative records indicating prior outreach modality (e.g., direct visit, mail-only contact, other)

These items are not mandatory, but they would be appreciated if they can be provided.

D6. Consultation Log (D) [Required]

Basic Structure

- One row per consultation
- One variable per column
- Linked to the household master by the administrative management number

- If the administrative management number is unknown, the cell may be left blank; if possible, supplementary location information such as aza ID and administrative district ID should be recorded

Main Item Names / Descriptions / Format (Examples)

Item	Description	Format (example)
Consultation ID	Sequential number within the consultation log	Numeric or text
Administrative management number	Link key to household master	Numeric or text
Consultation date	Date of contact / consultation	Date
Contact channel	Phone, counter/in person, visit, other	Code
Staff ID	Anonymized staff code	Text
Consultant category	Self, co-resident family member, non-co-resident family member, neighbor/other proxy, etc.	Code
Aza ID	Area identifier	Numeric or text
Administrative district ID	Area identifier	Numeric or text
Sewer service area classification	Inside area, outside area, mixed, unknown	Code
Current treatment status	Combined septic tank, single septic tank, night-soil collection, unknown	Code
Reason for consultation / trigger	Flyer, newsletter article, family member, neighbor, etc.	Code
Information heard	Information received via family, neighbor, etc.	Code
Saw the flyer?	Saw it, did not see it, do not remember, etc.	Code
Main concern before consultation	Cost, construction burden, unclear procedures, etc.	Code
Consultation content	Multiple selection allowed	Checkbox columns (recommended) or text
Consultation level	Level 1, 2, 3 (staff judgment)	Code
Contacted contractor	Already contacted, will contact, undecided	Code
Next step	A–E, etc. (staff judgment)	Code
Concern status after consultation	Resolved, somewhat remains, unchanged, etc.	Code

Item	Description	Format (example)
Clarity of pathway	5-point scale, etc.	Code
Research exclusion	0/1	Numeric

Operational Principles for the Consultation Log

- “Consultation content” may involve multiple applicable items; multiple selections are assumed.
- To prevent inconsistent wording, code input via pull-down lists (data validation) is recommended for key items.
- “Not confirmed” and “No response” should be included as options where appropriate in order to reduce blank cells.

Research exclusion should be coded as 1 when a resident indicates that they do not wish their information to be used for research. The ordinary administrative record itself should still be retained as needed.

D7. Supplementary Time-Use Record for Consultations [If feasible]

As a supplementary measure of implementation cost, the City may be asked, during a limited observation window, to maintain a simple record of the time spent handling telephone or in-person consultations.

This supplementary record is separate from the main consultation log and is intended only to provide a descriptive measure of the approximate administrative workload associated with the intervention. It is not intended to serve as a detailed work log, attendance record, or staff performance record.

The record may be maintained either in a separate file or in a separate worksheet within the same Excel workbook used for the consultation log.

As currently planned, this supplementary time-use record will be maintained only during limited observation windows, namely from the municipal newsletter delivery date (typically the 25th of the month) through the end of the first week of the following month, for the relevant months during the implementation period. The exact window will be finalized in coordination with the municipality before implementation.

Suggested Variables

The following variables are suggested for the supplementary time-use record:

- No.
- Date of record

- Approximate consultation start time
- Approximate consultation end time
- Approximate consultation duration
- Handling category (simple / standard / complex)
- Administrative management number (if identifiable)
- Consultation mode (if management number is unknown): telephone / counter
- Consultation type (if management number is unknown): sewer / septic tank / both / unknown
- Notes

Suggested Recording Rules

If the administrative management number is identifiable, the consultation mode and consultation type may be left blank.

If the administrative management number is not identifiable, the consultation mode and consultation type should be recorded where possible.

The handling category may be selected based on staff judgment using the categories *simple*, *standard*, and *complex*.

The observation window for this supplementary record will be determined in advance in coordination with the researcher.

Suggested Pull-Down Options

Suggested pull-down options are as follows:

- Handling category: simple / standard / complex
- Consultation mode: telephone / counter
- Consultation type: sewer / septic tank / both / unknown

D8. Data Delivery Format

The researcher would appreciate receiving the data in the following format.

(1) Household Master

- One row per household
- One variable per column
- Primary key: administrative management number

(2) Consultation Log

- One row per consultation
- One variable per column
- Linked to the household master by the administrative management number

Recommended Format

It is recommended that the data be created within a single Excel file, separated into:

- `Household Master` sheet
- `Consultation Log` sheet

If this is difficult in practice, separate files are also acceptable. If possible, please also attach a codebook or data dictionary explaining each column.

D9. Anonymization Procedures

- Within the City administration, create research-use data with personal information (names, addresses, etc.) removed.
- Use the administrative management number to identify households.
- Researchers will not receive names, addresses, or other personal information.
- The correspondence table linking management numbers to personally identifying information will be retained by the City, and researchers will not have access to it.
- The consultation log will be linked to the household master using the administrative management number.
- Staff will be recorded using anonymous staff IDs, and staff names will not be provided to researchers.
- If a resident requests that their data not be used for research, the administrative record will still be retained as needed for routine purposes, while a research exclusion flag will allow the corresponding observation to be removed from analysis.
- Researchers will not attempt to re-identify any individual or household.

D10. Data Entry Rules (Recommended)

For many consultation-log items (contact channel, trigger of consultation, main concern, next step, etc.), code entry via pull-down lists (data validation) is recommended to prevent inconsistent wording.

In particular, it is desirable to explicitly distinguish:

- **Not confirmed**: when there was not enough time to ask
- **No response**: when the resident did not wish to answer or did not answer

This makes it possible to distinguish such cases analytically from simple blank cells.

D11. Update Frequency

Updated data during the RCT period are expected to be provided monthly, with each submission closed at the end of the month, beginning in July 2026.

As needed, the researcher will review the received data and may contact the City if clear omissions or inconsistencies are identified.

D12. Notes

This document is intended to be read together with the main PAP and Appendix E (Consultation Log Items and Operational Notes).

This document summarizes the requested data items and delivery format for the sewer connection encouragement RCT. The exact column names and details of the data validation rules may be adjusted to fit administrative practice. In such cases, it is desirable to consult with the researcher and remain as consistent as possible with the principles of this specification.

E Front-desk Log Items and Operational Notes

Front-Desk Consultation Log Items and Operational Notes
For the Sewer Connection Encouragement RCT only[4]
Prepared as an appendix to the PAP

E1. Purpose of This Appendix

This appendix specifies the front-desk (consultation) log items to be recorded when residents make inquiries or consultations related to sewer connection encouragement in Date City. Its main purpose is to structure consultation information recorded in the course of ordinary administrative work so that it can be used for policy improvement and research analysis.

Researchers will not receive personally identifiable information such as names or addresses. This appendix should be read together with Section 5.3 of the PAP, the subsequent analysis plan, and Appendix D (Data Provision Request Specifications). It specifies both the log items and the operational rules for using the consultation log.

E2. Front-desk Log Items

1. Basic Information

- Consultation date
- Contact channel: telephone / counter visit / visit by city staff / other
- Staff ID (anonymous code)
- Consultant category: self (resident) / co-residing family member / non-co-residing family member / proxy such as neighbor / other / unknown

2. Household Identification and Area Information

- Administrative management number (if available)
- Aza ID
- Administrative district ID (if available)
- Sewer service area category: inside service area / outside service area / mixed / unknown
- Current wastewater treatment status: combined septic tank / single-treatment septic tank / pit latrine / unknown

[4] This appendix is a research-oriented English version prepared for the PAP. This appendix is prepared for the Sewer Connection Encouragement RCT only. Although some front-desk operations may, in practice, overlap with other wastewater-policy consultations, the present appendix defines the log items and operational notes required for the sewer connection encouragement RCT.

3. Trigger for Consultation and Information Reach

- Trigger for consultation: city newsletter flyer (insert) / city newsletter article (other than the flyer) / co-residing family member / non-co-residing family member / neighbor outside the household / contractor or other external party / other / no response / not confirmed
- If the trigger was family or neighbors: what the resident heard (consultation contact information / costs or construction / personal experience / that they saw the city notice / other / no response / not confirmed)
- Whether the resident saw the flyer: saw it / did not see it / does not remember / no response / not confirmed
- If seen: how the flyer was obtained (newsletter insert / already at home / from family or acquaintances / from contractor or other source / does not remember / no response / not confirmed)
- If obtained through the newsletter insert: which issue (July / October / November / unknown / no response / not confirmed)

4. Main Concern Before Consultation

- Main concern before consultation: cost / burden of construction (time and effort) / procedures unclear / how to choose a contractor / housing conditions such as indoor plumbing work / other / no response / not confirmed

5. Consultation Topics (Multiple Responses Allowed)

When multiple topics apply, staff should record all relevant topics.

- Confirmation of whether the property is in the service area
- Connection method
- Cost
- Introduction to contractors
- Coordination with other works
- Application guidance
- Other

6. Consultation Level (Staff Judgment)

- Consultation level: Level 1 (information provision) / Level 2 (specific consideration) / Level 3 (premised on implementation)

  This should be recorded based on staff judgment rather than resident self-assessment.

7. Status of Contact with Contractors

- Whether the resident has contacted a contractor: already contacted / will contact / undecided / not confirmed
- If already contacted: timing relative to the consultation with the City (before City consultation / after City consultation / unknown / not confirmed)

8. Next Step (Staff Judgment)

- Next step: A (completed on the spot) / B (materials to be sent) / C (follow-up contact scheduled) / D (additional City response) / E (procedural guidance) / other / not confirmed

  This should also be recorded based on staff judgment rather than resident self-assessment.

9. Optional Confirmation Items

9-1. Clarity of the Action Pathway (Optional)

- Very easy to understand / somewhat easy to understand / neither / somewhat difficult to understand / very difficult to understand / no response / not confirmed

9-2. Anxiety Status After the Consultation (Optional)

- Resolved / somewhat remains / unchanged / do not know / no response / not confirmed

10. Research Exclusion (Opt-out)

- Research exclusion flag: 0 (eligible for research use) / 1 (exclude from research)

If a resident states that they do not wish their record to be used for research, ordinary administrative records should still be retained as needed, but the research exclusion flag should be set to 1. This flag has no effect on administrative handling and is used only to exclude the case from research analysis.

E3. Operational Notes

1. Priority to Ordinary Administrative Work

Ordinary administrative work should always come first. This log is intended for use only within the range that can be recorded naturally in the course of ordinary work. Staff should prioritize resident support rather than attempting to complete every field.

2. Prefer "Not Confirmed" or "No Response" to Leaving a Blank

Use **not confirmed** when there was not enough time to confirm an item. Use **no response** when the resident did not wish to answer or did not answer. These options are preferable to leaving cells blank.

3. Take Care When Speaking with Older Residents

Avoid technical terms where possible, use concise and clear expressions, and do not force every item into the conversation.

4. Multiple Consultation Topics May Apply

Consultation topics allow multiple selections. When multiple topics apply, staff should record all relevant items.

5. Consultation Level and Next Step Are Based on Staff Judgment

The consultation level and next step should be recorded based on staff judgment, not resident self-assessment.

6. Keep a Record Even If the Management Number Is Unknown

Even if the administrative management number is unknown, staff should still keep the consultation record. If possible, aza ID and administrative district ID should also be recorded.

7. Research Exclusion Must Not Affect Administrative Services

If a resident does not wish their information to be used for research, add the research exclusion flag. This must not affect consultation, procedures, or access to administrative services.

8. Separate Supplementary Time-Use Record

If a simplified consultation time-use record is maintained, it should be kept separately from the main consultation log. The main consultation log should always take priority, and the supplementary time-use record should be completed only within the designated observation windows and to the extent feasible.

E4. Suggested Explanations If Residents Ask About the Research

Q1. If Asked, "Is this research?"
This is part of an effort to make the City's notices and consultation services easier to understand and improve them over time. Records may be used for policy improvement and analysis only after they have been handled in a form that does not identify individuals.

Q2. If Asked, "Do I Have to Answer?"
No. It is fine not to answer any item that the resident does not wish to answer. This does not affect procedures or access to administrative services.

Q3. If Asked, "Please Do Not Use This for Research"

Understood. The City will retain the record as necessary for ordinary administrative work, but the case will be excluded from research analysis.

E5. Related Q&A

Q1. Do staff have to ask every item?

A. No. Ordinary administrative work should come first, and staff should record items only within a reasonable range.

Q2. What if the management number is unknown?

A. Staff should still keep the consultation record. If possible, they should also record the aza ID and administrative district ID.

Q3. What if the resident consults multiple times?

A. Record each consultation as one row per consultation.

Q4. What if multiple consultation topics apply?

A. Select all applicable topics.

Q5. What if the resident says their information should not be used for research?

A. Retain the record as necessary for ordinary administrative work and add the research exclusion flag.

E6. Note

This appendix provides an overview of the consultation-log items and their operational use for the sewer connection encouragement RCT. Detailed column names, validation rules, and code lists are specified separately in Appendix D (Data Provision Request Specifications).

F Research Timeline

Research Timeline
For the Sewer Connection Encouragement RCT only[5]
Prepared as an appendix to the PAP

At the current stage, the sewer connection encouragement RCT follows the timeline below. The schedule reflects the current implementation plan agreed upon with Date City. Minor adjustments may be made depending on data availability and administrative coordination, without affecting the core research design or identification strategy.

2025

Month	Activities
December	• Receipt of the first dataset (latest version available in 2025) • Preliminary checking of household-level data structure and variable definitions • Development of the preliminary cluster design • Preparation of the preliminary pre-analysis plan (pre-PAP)

2026

Month	Activities
February	• First round of coordination with Date City regarding the sewer connection encouragement leaflet, consultation-log items, and possible follow-up survey • Finalization of the leaflet content and layout
March	• Submission of the ethics application by the researcher • Revision of the PAP, related appendices, and related administrative documents based on this coordination

[5] This appendix is a research-oriented English version prepared for the PAP. This appendix is prepared for the sewer connection encouragement RCT only. Although some administrative recording procedures may in practice be shared with the combined septic tank conversion promotion RCT, the present document specifies the research timeline for the sewer RCT and should be interpreted accordingly.

Month	Activities
April	• Revision of the PAP, related appendices • Baseline operation of consultation-log recording and related administrative preparation • Administrative updating of household records to reflect moves and related changes
April–June	• Baseline observation windows for simplified consultation time-use recording, if implemented
Early May	• Receipt of the frozen dataset for final treatment assignment • Data consistency checks and final verification of the randomization procedure
Mid May	• Final cluster-level treatment assignment (one-time assignment based on the frozen dataset) • Finalization of the full PAP • Registration with the AEA RCT Registry
Late May–June	• Printing and administrative preparation for leaflet distribution • Finalization of consultation-log operation and staff briefing on recording procedures • Preparation for the collection of implementation records and administrative cost records, where feasible
July	• Start of the first intervention phase (distribution of the finalized leaflet to the first-wave clusters through the municipal newsletter) • Start of monthly submission of updated administrative records and consultation logs • Start of the first post-distribution observation window for simplified consultation time-use recording and administrative workload measurement, if implemented

Month	Activities
August–September	• Ongoing collection of administrative records and consultation logs • Ongoing review of implementation fidelity and data consistency
October	• Start of the second intervention phase (distribution of the finalized leaflet to the second-wave clusters) • Start of the second post-distribution observation window for simplified consultation time-use recording and administrative workload measurement, if implemented
October–November	• Continued collection of administrative records and consultation logs • Continued review of short-run implementation and consultation activity
December	• Follow-up survey of households with no consultation record, if implemented • Preliminary review of short-run outcomes

2027

Month	Activities
January–March	• Continued collection of administrative outcome data • Data cleaning and construction of the analysis dataset
April–June	• Main empirical analysis
July–December	• Additional data collection for medium-run outcomes, if available • Robustness checks, manuscript preparation, and conference presentations

Notes

1. The final treatment assignment will be conducted only once, using the frozen household dataset provided in early May 2026.
2. Pre-assignment checks conducted before May 2026 are intended only for data consistency verification and trial runs of the randomization procedure; they do not constitute the formal assignment.
3. The primary intervention consists of distributing the finalized information leaflet through the municipal newsletter in two waves, in July and October 2026.
4. The consultation log is part of ordinary administrative workflow and is used to document consultation-level information in a form suitable for research analysis.
5. The follow-up survey, if implemented, is intended as a supplementary mechanism study for households with no consultation record and is not part of the primary outcome measurement.
6. The current plan is to receive administrative outcome data through March 2027; if feasible, data provision may be extended to cover later application and completion outcomes through December 2027.
7. Some operational procedures may in practice overlap with those used for other wastewater-policy activities. However, the timeline above is presented specifically for the sewer connection encouragement RCT.

G Draft Follow-Up Survey Plan

Draft Follow-Up Survey Plan
For the Sewer Connection Encouragement RCT only[6]

G1. Purpose of This Appendix

This appendix outlines the current plan for the follow-up survey to be conducted as a supplementary module for mechanism analysis in the Sewer Connection Encouragement RCT.

The follow-up survey is not part of the primary outcome measurement and will not be used to redefine the main analysis sample or the primary outcome. Rather, it is intended to provide supplementary evidence on mechanisms among households that do not appear in the consultation logs, especially with respect to leaflet recognition, understanding of the consultation pathway, beliefs about eligibility, and reasons for not consulting.

The present appendix describes the planned target population, timing, sampling approach, division of responsibilities, and main measurement domains. The detailed questionnaire wording may be finalized later and attached in a subsequent supplementary version if needed.

G2. Role of the Follow-Up Survey in the Study

The main analysis of the RCT relies on administrative behavioral history data and consultation logs. These sources are well suited for identifying whether households consulted, requested estimates, applied, or completed connection-related procedures. However, they do not directly reveal why some households did not consult at all.

The follow-up survey is therefore intended to complement the administrative records by collecting information from households with no consultation record during the relevant evaluation period. In particular, it is intended to help distinguish among several possible mechanisms, including:

- the leaflet did not reach or was not noticed by the household,
- the household did not understand where to seek consultation,
- the household believed that it was not in the target area,
- the household faced financial, procedural, or other decision-making constraints, or
- the household had no current intention to consider sewer connection.

The follow-up survey is thus treated as a *supplementary module for mechanism analysis*, rather than as a main source for causal identification.

[6] This appendix is a research-oriented English version prepared for the PAP.

G3. Target Population

The target population for the follow-up survey consists primarily of households with no consultation record by the end of November 2026.

In principle, the survey will focus on non-consulting households that remain in the study population after applying the relevant research-exclusion rules. Households that have opted out of research use will not be included in the survey sample for research purposes.

G4. Planned Timing

The current plan is to conduct the follow-up survey in December 2026, after the main short-run evaluation period.

This timing is intended to allow:

- sufficient time for the first-wave and second-wave leaflet distributions to generate consultation responses,
- identification of households with no consultation record by the end of November 2026, and
- implementation of the survey while short-run recall remains feasible.

Minor adjustments to the exact mailing date may be made depending on administrative coordination and operational constraints.

G5. Sampling Plan

The follow-up survey sample will be selected by stratified random sampling.

The current plan is to ensure balanced representation from:

- the first-wave clusters, and
- the second-wave clusters.

If needed, mixed-area aza units may be oversampled in order to improve the ability to study misunderstanding about service-area eligibility and related information frictions.

The detailed sampling rule, including the final stratification scheme, random seed, and reserve sample procedure, will be fixed in advance before sample selection is conducted.

G6. Planned Sample Size

The planned mailing scale is approximately 600 households, subject to administrative feasibility, budget constraints, and the expected response rate.

At the current stage, this sample size is intended primarily to provide informative descriptive evidence on mechanism-related outcomes rather than precise causal estimates. The final sample size may be adjusted in consultation with the City depending on mailing costs, operational burden, and expected response rates.

G7. Main Measurement Domains

The follow-up survey is expected to collect information on the following domains:

1. **Leaflet recognition**
   Whether the household recalls receiving or seeing the leaflet.

2. **Understanding of the consultation pathway**
   Whether the household knows where to seek consultation and what can be asked through consultation.

3. **Recognition of target-area eligibility**
   Whether the household believes that it is located in an area relevant to sewer connection encouragement.

4. **Reasons for not consulting**
   Main reasons for not contacting the City or a contractor, including cost concerns, procedural uncertainty, lack of perceived necessity, uncertainty about eligibility, or other reasons.

5. **Decision-making constraints**
   Constraints related to cost, procedures, timing, family coordination, uncertainty, or other barriers to taking the first step.

6. **Future consultation intention**
   Whether the household intends to seek consultation in the future.

The exact question wording, response scales, and ordering may be refined later. The detailed questionnaire form may be attached separately once finalized.

G8. Planned Division of Responsibilities

The anticipated division of responsibilities is as follows:

- The researcher will determine the sampling rule, conduct sample selection using management-number-based data, and provide the City with a list of sampled management numbers.
- The City will attach names and addresses to the sampled management numbers, prepare mailing materials, send the questionnaire, and receive returned responses.
- The researcher will not receive names, addresses, or other personally identifying information.

If reserve samples are used, the reserve sample list will also be prepared by the researcher using the same pre-specified randomization procedure.

G9. Personal Information Protection and Research Ethics

The follow-up survey will be implemented in a manner consistent with the overall research ethics framework of the study.

- The researcher will not receive personally identifying information such as names or addresses.
- The correspondence table linking sampled management numbers to names and addresses will be retained by the City.
- Participation in the follow-up survey will be voluntary.
- Whether or not a household responds to the survey will not affect consultation, application, or any other administrative procedures.
- If a household indicates that it does not wish its responses to be used for research, the relevant research-exclusion rule will be applied.

G10. Analytical Role and Limitations

The follow-up survey will not be used to redefine the main analysis sample or the primary outcome, and it will not be treated as the main source for causal identification.

Because survey participation is voluntary and nonresponse may be non-random, results from the follow-up survey will be interpreted as supplementary evidence for mechanism analysis rather than as the primary basis for causal inference.

Where appropriate, the study will report response rates and compare respondents and nonrespondents using available administrative characteristics.

G11. Current Status and Future Revision

At the current stage, the broad design of the follow-up survey has been discussed and is broadly acceptable in principle. The main remaining tasks are:

- finalizing the detailed questionnaire wording,
- confirming the mailing workflow with the City,
- confirming the final sampling rule and reserve-sample procedure, and
- finalizing the schedule for mailing and response collection.

If necessary, a later supplementary version may attach the finalized questionnaire, mailing materials, or additional implementation details.