Code-book

Search Process

Meta analyses need three levels of information to be coded: search level, study level and the effect level information. We have done the first two by using the apsis platform (https://arxiv.org/abs/2405.04621) which automatically records our search strategy and study level information from the documents that have been tagged as relevant. This document mostly relates to the manual coding at the effect level, for which we used excel.

Ineligible documents

After going through the abstract and reading through the document, if you find it unacceptable to be included in the study because 1) it does not contain relevant effect sizes 2) effect size information is incomplete or irretrievable 3) document is not relevant or any other reason, please add the reason for excluding the document exclusion reason field.

Acceptable documents

* For acceptable documents we collect information on the “Effects” and the “Interventions”.
* Each study may report multiple estimates of the effect. This can be due to multiple iterations, multiple treatments/interventions, different populations, etc. You can add the multiple effects and corresponding information for each study. As such, please collect information on all the effect sizes estimated in the study along with the corresponding control variables.

Effect Fields

In this section we will capture one estimation (e.g. one coefficient, one difference of means) and the characteristics of the model used in order to accurately calculate effect sizes comparable to others included in our meta-analysis.

After the statistical information is captured it will be linked to an intervention in the next section. An effect size can be estimated for joint interventions (e.g. a TOU pricing scheme might be introduced alongside an in-home display device). The capture of the intervention section should describe the intervention linked to the estimate as accurately as possible.

| **Field Name** | **Explanation** | **Choices or Examples** |
| --- | --- | --- |
| Document Type  *Variable:*  *DocumentType* | Code the type of document:  **Peer-Reviewed Paper,** if the document is an academic paper, published in a peer-reviewed journal.  **Unpublished Academic Paper,** if the document is an academic paper which has not been published in a peer-reviewed journal, e.g. a conference/working/discussion paper. | Peer-Reviewed Paper  Unpublished Academic Paper |

|  |  |  |
| --- | --- | --- |
| Page  *Variable:*  *Page* | Capture page # for the specific effect – preferably the page number that presents the table of results.  If you only have a version of the paper without page numbers use a normal page count to fill in the relevant page number. |  |
| Effect number  *Variable:*  *EffectNumber* | Assign numbers to the effect making sure they are identifiable at a later stage. In general, numbers should get larger for effects which appear later in the text. | Attention: effect numbers do not always start with a 1. There is no particular meaning to the effect numbers and missing numbers do not imply missing effect sizes. |
| Statistical technique  *Variable:*  *statistical\_technique* | Studies may employ different techniques to estimate effects.  Note: Don’t code Difference in Difference as effect statistical\_technique. Here we really just capture the type of regression and let Difference in Difference be the study type | NA,  Difference of means,  ANOVA,  One-step GMM,  Household fixed effects,  Time fixed effects,  Household and time fixed effects,  Random effects regression,  OLS. |
| Dependent variable  *Variable:*  *dependent\_variable* | Studies will utilize various measures to capture energy consumption. What variable/operationalization are the authors using here. We are ultimately interested in percentage change in energy consumption of the household given the intervention.  Note:  We still capture experiments where households own automated (smart) thermostats, as long as the settings of the thermostat can be overridden at any time (eg. Herter (2013))  We always capture the ITT and disregard ATT/LATE estimates.  Write the dependent variable and the unit between brackets | Example: household hourly electricity consumption (kWh) |
| Study design  *Variable:*  *study\_design* | Whether the study calculates a pre-test/ post-test effect size or a control-treatment effect size  *Pre-test/ post-test*- No control used, same group measured twice  *Control-treatment* - no baseline, two groups with starting observation started at the same time  *Difference in Difference* - baseline for both control and treatment groups followed by treatment and measure of effect | NA,  Pre-test/ post-test,  Control-treatment,  Difference in Difference |
| Control  *Variable:*  *control* | Describe the control group of the study – whether the treatment is being compared to absence of treatment or a different treatment configuration. | No treatment,  Any available treatment configuration that serves as a control group in this study (values are defined by intervention types and sub-types) |
| ­Effect size - statistical estimate (coefficient)  *Variable:*  *coefficient* | Capture the *absolute* value of the relevant regression coefficient. |  |
| Baseline consumption  *Variables:*  *Baseline Consumption*  *Notes\_baselineConsumption* | Record the average consumption in energy units (eg. kW/h) during the period which the coefficient/difference in means refers to.  We want use this variable to calculate the percentage change in energy consumption during the analyzed period when the coefficient/ difference in means must be interpreted as the change in energy units (dividing the coefficient/ difference in mean by the baseline consumption).  Rules for recording the baseline consumption:  **Which consumption?** If baseline data was collected and average consumption during that time is reported, record that. If baseline is not available, record the average consumption during the experiment.  **Whose consumption?** If average consumption of all households is mentioned, record that. If not, then the weighted average consumption of treatment and control group should be reported. If both the options are not possible, report the average consumption of the control group.  Preference order:   1. Average consumption of all households (T+C) during baseline period 2. Weighted average consumption of T and C during baseline period 3. Average consumption of all households (T+C) during the treatment period 4. Weighted average consumption of T and C during treatment period 5. Consumption of the control group |  |
| Direction of effect size  *Variables:*  *effect\_direction* | Capture the direction of the effect of the intervention. From the baseline review 2024 onwards, we capture effect size direction as a descriptive understanding of the effect of the intervention.  Example:  *Increase*, if energy consumption increased after exposure to intervention (undesired effect). | NA  Increase  Decrease  No change |
| Coefficient SD  *Variables:*  *coefficient\_sd* | The variance is captured by the standard error of the coefficient. |  |
| Coefficient SD type  *Variables:*  *coefficient\_sd\_type* | Also capture the type of uncertainty measure provided (see right). | NA,  Standard error,  Standard deviation,  Robust standard errors,  Clustered standard errors,  Bootstrapped clustered standard errors,  Pooled standard errors,  Corrected variance matrix,  Not provided |
| Missing Uncertainty measures  *Variable :*  *uncertaintyMeasuresMissing* | Indicate if uncertainty measures are missing but sample size is provided.  Leave blank if uncertainty measures are provided. | No uncertainty measures only sample size available |
| Significance  *Variables:*  *p\_value*  *significance\_bound*  *significance\_test*  *test\_statistic*  *effect test\_tails* | Also capture the t-statistic.  If a study provides the coefficient and standard error, calculate the specific t-statistic (β/se) and use that to calculate a corresponding P-value.  If a study only provides the level of significance (no standard error, no t-statistic, no specific p-value) then capture the level of significance in the p-value field and calculate the t-statistic using the level of significance and mark it as a lower bound (if significant) or choose the p-value 0.1 and upper bound (if not significant)  [**https://handbook-5-1.cochrane.org/chapter\_7/7\_7\_3\_3\_obtaining\_standard\_deviations\_from\_standard\_errors.htm**](https://handbook-5-1.cochrane.org/chapter_7/7_7_3_3_obtaining_standard_deviations_from_standard_errors.htm)  Example:  *Upper bound*, if the effect is significant at p-value=0.05 but a coefficient SD is not provided.  *Lower bound*, if the effect is not significant and the significance test is for p-value=0.1 but a coefficient SD is not provided. | *NA*  *lower bound*  *upper bound*  *actual* |
| Effect size - difference of means  *Variables:*  *control\_mean*  *control\_sd*  *treated\_mean*  *treated\_sd*  *diff\_mean*  pooled\_sd | Capture both the control and treatment group means where possible. Otherwise capture the difference along with the value of the relevant test statistic (t, Chi or F).  Calculate pooled standard deviation using the formula in Ringquist if required. | If the control sample size is not clearly specified, take a conservative approach and use the treatment sample size as total sample size. |
| Sample Size  *Variables:*  *control\_sample\_size*  treatment\_sample\_size  total\_sample\_size | Capture as much detail as possible (i.e. if all three options are given, record all options indicated below). Given that a specification can include multiple treatments, the sum of treatment and control sample sizes will not necessarily add to the total.  Studies may run analysis comparing treatment groups only to the control, or also to each other, capture elements    **Total** – Control and all relevant treatments, full sample size (for pre-and post- treatment set-ups)  **Treatment** – Capture sample size for specific effect being captured (this should correspond to one treatment or combination of treatments)  **Control** – There is possibly only one control group for multiple treatments |  |
| Control Definition  *Variables:*  *controls Electricity use*  *controls Energy prices*  *controls Environmental attitudes*  *controls HH controls (demographics)*  *controls None - base model*  *controls Residence controls*  *controls Seasonal controls*  *controls Weather controls* | List the controls which are being used to better isolate the effect of the intervention. Besides the variable of interest (intervention) which elements that affect energy consumption are included.    Fixed effects and random effects for households or time effects should NOT be captured here. These should be captured in the statistical technique. Only include explicit weather or seasonal controls in the analysis.  If in pre/post analysis the baseline is corrected for weather, code 1 for weather controls. | **Weather controls** (heating degree days or cooling degree days, etc.)  **Seasonal controls** (monthly dummies or quarterly/seasonal dummies, etc)  **Energy prices**  **Residence controls** (physical nature of the house, size of house, appliance stock, etc.)  **Household controls** – demographic info (income, age, number of residents, education, etc.)  **Base energy consumption** |
| Geography and Aggregation Level  *Variables:*  *geographic\_scope*  geographic\_location  *Notes: Study scope*  *City*  *State*  *country* | **Geographic scope** captures the area across which the intervention was carried out  Note: Try and stick to the options mentioned  Leave a note if smaller than municipality.  The field **Geographic location** should not be codedanymore. Instead, use the separate fields city, state, country.  Note: Try and stick to the format city, state, country | For Geographic scope:  smaller than municipality, municipality, state, town, county |
| Smart metering  *Variable:*  *smart\_metering* | If smart metering was used at which aggregation level was the energy consumption metered.  Choose “not specified” if smart metering was used but the aggregation level is not reported.  Do not fill out. | Yes/No  NA |
| Randomization  *Variables:*  *randomisation\_dummy*  *randomization\_method* | Were households randomly assigned to control and treatment group and/or between different treatment groups?  Were households in a given study randomized at the cluster level (district, state, neighborhood) or at the household level? | Yes/No  For randomization method:  Cluster, household |
| Opt-in vs. Opt-outs  *Variable:*  *Opt\_in* | Were households first selected and then allowed to opt out of the intervention or were the households required to opt-in to the intervention or neither  Note: Write “Opt in” when households are given the option to opt-into the experiment explicitly. Choose “Opt out” when households are given an option to drop out of the experiment explicitly.  Choose “Mandatory” if the participation in the scheme was mandatory with no option to opt-out. | Opt in  Opt out  Mandatory  Neither  Unclear |

Intervention

| **Field Name** | **Explanation** | **Choices or Examples** |
| --- | --- | --- |
| Total Duration  *Variable:*  *interventionDuration* | Capture the *total* duration of the experiment.  Note: Data for duration should be in weeks where 1 month = 4 weeks and 1 year = 52 weeks | Pre/post setting: experiment runs for 3 months, control period is the same 3 months in the year before -> 12  An experiment runs for two subsequent summers. The intervention duration in the first summer is 12 weeks, in the second summer the intervention duration is 64 weeks.  interventionDuration is the total duration of the study including the baseline observation period, treatment period and follow up period if any. |
| Treatment Duration  *Variable: InterventionTreatmentPeriod* | For how long did the *treatment* of the intervention take place.  Note: Data for duration should be in weeks where 1 month = 4 weeks and 1 year = 52 weeks | If two separate periods are pooled together, the intervention duration is the sum of both intervention treatment periods. |
| Follow-up dummy  *Variable:*  *interventionFollowup* | Record yes if a follow-up investigation into energy consumption was conducted and no otherwise. | Yes  No  Unclear |
| Framing Unit  *Variable:*  *framing\_unit* | When participants are given information or feedback on their energy consumption, in what units/terms is this information given | Energy (kwh)  savings ($)  Co2  other efficiency (R-value)  other ecological |
| Timing/Frequency  *Variable:*  *informationFrequency* | What is the frequency at which the households are contacted to provide feedback/ information? Normally not applicable for pricing interventions. | Continuous (on-demand)  Monthly  Bi-weekly  One time only  Monthly or quarterly |
| Medium  *Variable:*  *medium* | How was the household contacted? Could be relevant for all types of interventions. |  |
| Intervention types and sub-types | We look at five broad intervention types: Information, Feedback, Monetary Incentives, Social Comparison and Motivation | Information: home audits, tips,  Feedback: historical, enhanced billing, in-home display  Social comparison: peer comparison, HER  Monetary Incentives: Rewards, ToU, Real-time Pricing, Dynamic Pricing, Peak Load Pricing  Motivation: Gamification, commitments, goal setting |
| Exclude effect size  *Variables:*  *EffectSizeExclusion*  *EffectSizeExlusionReason* | If an effect size was already coded, but you thereafter realize that it can not be included e.g. because of missing uncertainty measures choose “exclude” and describe the reason which lead to this decision. | *Exclude* |

Risk of bias assessment (CEE CAT)

|  |  |  |
| --- | --- | --- |
| ConfounderPossibility [Criterion 1, 1.1, pg. 16] | Is it possible for the effectiveness of the intervention to be confounded? (i.e. could there be a bias due to uncontrolled or inappropriately controlled variable that influences both the intervention and the outcome) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| ConfounderAll [Criterion 1, 1.2, pg. 16] | Did the authors control for all the potential confounders? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| ConfounderJustifiedOmission [Criterion 1, 1.3, pg. 16] | Is there any justifiable reason for not controlling for all the potential confounders (so that omission of some of the potential confounders is unlikely to influence the assessment of the effectiveness or impact)?  (E.g., select Y/SY when there is evidence that omission of some of the potential confounders does not affect the assessment of effectiveness or impact. This may be the case if adjusting all potential confounders will lead to overadjustment, or an ‘instrumental variable’ is used for estimating the effectiveness or impact, etc. Instrumental variable is a variable that (1) is not associated with the confounder(s), (2) is associated with the intervention/exposure but (3) does not directly influence the outcome. If used appropriately, it enables valid estimation. See Hernán & Robins 2020 for guidance.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| ConfounderAccuracy [Criterion 1, 1.4, pg. 16] | Were the potential confounders, that were controlled for, (and/or the instrumental variable used if applicable) likely to be measured accurately and precisely enough?  (Measurements of factors may be nominal (categorical), ordinal (ranks) or scale.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| ConfounderAnalysis [Criterion 1, 1.5, pg. 16] | Did the author(s) analyse the effect appropriately by taking into account the potential confounders, as well as the issue of accuracy and precision of the measurements of the potential confounders (and the instrumental variable if applicable)?  (Examples of appropriate adjustment techniques for confounding may include stratification, matching, inverse probability weighting, standardisation, G-estimation, and instrumental variable estimation.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| SampleExchangeability [Criterion 2, 2.1, pg. 17] | Was the selection of subjects or areas after intervention or exposure random or systematic (i.e., based on random or systematic sampling), and exchangeability between groups could be assumed based on the selection approach? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| SampleExclusion [Criterion 2, 2.3, pg. 17] | After the start of the intervention/exposure or during the analysis, were any subjects or areas excluded or lost from the study or analysis?  (When some subjects or areas, or collected data are excluded, it might increase the risk of post-intervention/exposure selection bias.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| SampleGroupComparability [Criterion 2, 2.4, pg. 17] | Were the subjects or areas included in the study (or analysis) comparable between groups and so they allowed a valid comparison to be made (i.e., exchangeability or conditional exchangeability between groups could be assumed)? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| SampleGroupDifferenceIntervention [Criterion 2, 2.5, pg. 17] | Were the difference(s) between groups likely to be explained by the intervention/exposure or a variable influenced by the intervention/exposure (including the outcome)? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| SampleBiasAdjustment [Criterion 2, 2.6, pg. 17] | Did the author(s) adjust for the potential post- intervention/exposure selection bias in an appropriate way? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| AwareofStudy [Criterion 5, 5.1, pg. 24] | Was there any way for the outcome measure to be affected by knowledge of the exposure, intervention, subjects or areas, or desire for certain outcome? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| DataSelective [Criterion 6, 6.1, pg. 26]  DataSelective | Are the reported effect estimate likely to only represent a part of measurements of the outcome?  (I.e., only a part of measured outcomes is reported. E.g., only 80 measured outcomes are reported when there are 100, or the effect estimate is based on 80 measured outcomes when there are 100.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| DataSubgroups [Criterion 6, 6.2, pg. 26]  DataSubgroups | Are relevant outcome data likely to be unreported for some subgroup(s)?  (I.e., only outcome data on certain subjects or areas with certain characteristic(s) (e.g., taxonomic group) or in certain conditions (e.g., intervention intensity) are available.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| DataCausal [Criterion 6, 6.3, pg. 27]  DataCausal | Is/are the analysis/analyses of the causal relationship of interest (intervention-outcome or exposure-outcome) likely to be partially reported?  (I.e., there is/are other relevant analysis/analyses of the causal relationship that is/are not reported.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| StatsRecording [Criterion 7, 7.1, pg. 28]  StatsRecording | Was/were the person(s), who estimated the effectiveness of the intervention or the impact of the exposure, aware of the exposure or intervention received by subjects or areas?  *No*, if the energy or electricity consumption are measured by meters rather than individuals who are aware of the intervention.  *Yes* otherwise. | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| StatsDescriptiveError [Criterion 7, 7.2, pg. 28] | Is it likely that there is/are error(s) or inappropriate methods in the applied descriptive statistical analyses? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| StatsInferentialError [Criterion 7, 7.3, pg. 28] | Is it likely that there is/are error(s) in the applied inferential statistics (including null hypothesis testing, estimation, coding)? | Yes  Seemingly Yes  No  Seemingly No  Not applicable |
| StatsInferentialViolation [Criterion 7, 7.4, pg. 28] | Were assumptions for the applied inferential statistics violated or the applied inferential statistical methods inappropriate for the inferential goal(s)?  (e.g. use of inappropriate sample sizes to test the hypothesis, normality not assumed when conducting a parametric test, equal or unequal variances not tested when testing for a difference, no justification for the choice of dependent and independent variables, a Pearson’s correlation test was used when analysing a causal relationship, inappropriate comparison of multiple models to support the provided statement when some of the models do not relate to impact or effectiveness, inappropriate modelling which may affect an estimate of effectiveness or impact.) | Yes  Seemingly Yes  No  Seemingly No  Not applicable |