HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Transporting behavioral insights to the poor: An RCT on energy efficiency investments among low-income households

Bettina Chlond

Timo Goeschl

Martin Kesternich

Madeline Werthschulte

AWI DISCUSSION PAPER SERIES NO. 755
September 2024

Transporting behavioral insights to the poor: An RCT on energy efficiency investments among low-income households

Bettina Chlond[1,2], Timo Goeschl[1,2], Martin Kesternich[3,2], and Madeline Werthschulte[2,4]

[1]Alfred Weber Institute for Economics, Heidelberg University
[2]ZEW – Leibniz Centre for European Economic Research
[3]Paderborn Research Center for Sustainable Economy, Paderborn University
[4]Institute for Environmental Studies, Vrije Universiteit Amsterdam

September 27, 2024

Abstract

OECD countries have recognized the urgent need to increase energy efficiency among low-income households. Which policy designs can be relied upon to effect this increase is an open question, but "behavioral insights" are likely components of future designs. Applying such insights to the economically disadvantaged raises problems of transportability: The studies at the heart of "behavioral insights" do not include sufficient data from the target population of low-income households to provide dependable estimates of causal effects. We illustrate the presence and scale of the transportability problem by conducting a randomized control trial on scalable, low-cost design elements to improve program take-up in one of the world's largest energy efficiency assistance programs. Observing investment decisions of over 1,800 low-income households in the Refrigerator Replacement Program, we find that the transportability problem is real and consequential: The most effective design would not have been chosen based on previous behavioral insights, and the design most in line with these insights backfires, violating 'do no harm' principles of policy advice. Systematic testing remains critical for addressing the transportability problem, particularly for policies targeting vulnerable groups.

JEL Classification: C93; D91; Q49

Keywords: Energy efficiency; field experiment; governmental welfare programs; nudges; program take-up

---

Notes: The experiment was pre-registered at the AEA RCT Registry as trial AEARCTR-0010757. This work has been funded by the Fritz-Thyssen-Foundation. Declaration of interests: None.

1 Introduction

Rapidly rising energy prices in many developed countries have highlighted to policy-makers the need to design and implement targeted policies for increasing energy efficiency among low-income households. While all households are negatively affected by higher expenditures for residential energy consumption, low-income households are particularly exposed. They already spend a large share of their disposable income on energy and their energy demand tends to be even less elastic compared to that of an average household (Schulte and Heindl, 2017). Many countries have recognized this urgent need. In the US, the Green and Resilient Retrofit Program funded under the Inflation Reduction Act has recently strengthened its effort to improve energy and water efficiency in low-income families (US Department of Housing and Urban Development, 2024). Similarly, in Europe, the EU Directive 2023/955 obligates Member States to submit plans to "prioritize energy efficiency improvements for vulnerable customers, low-income households, and individuals in social housing". As public funds become increasingly scarce, these plans are likely to feature increased attention on low-cost policies: Managers are looking for effective but low-cost policy options, making behavioral interventions ("nudges") an obvious candidate for future policy approaches.

Designing behaviorally-guided programs to achieve sizeable effects among a specific target group poses particular challenges for evidence-informed policy making. While policy-makers may be informed on "behavioral insights" (Halpern and Sanders, 2016; Gopalan and Pirog, 2017) emerging from an expanding set of carefully executed studies, that evidence base is often silent on the estimated effects of candidate policies on a specific subgroup like low-income households. The reason is that these households are typically absent or, at a minimum, systematically under-represented in untargeted programs. Consequently, applying these evidence-based insights to economically disadvantaged households raises concerns of "transportability": Can a policy-maker or program designer be confident that insights from specific interventions tested successfully elsewhere "will also work" for a new target group (Halpern and Sanders, 2016; Hallsworth, 2023)? Failure of transportability can lead to disappointing policy outcomes or, worse, to negative impacts on (vulnerable) target populations, possibly violating ethical rules of 'do no harm' in policy advice (Harrison et al., 2020).

This problem is common in evidence-based policy-making. Health policies, for example, increasingly rely on empirical evidence from Randomized Control Trials (RCTs). RCTs in this domain tend to be unrepresentative of the patient population (Goldstein et al., 2019) and often fail to feature patients with certain characteristics, such as ethnic minorities (Duma et al., 2018).[1] When

[1]The reasons for certain subgroups not being included can be both intentional and unintentional. Medical RCTs, for example, intentionally exclude patients with co-morbidities, existing prescriptions, drug abuse and unintentionally struggle to recruit from marginalized groups. Likewise in education research, it has become apparent that rigorously establishing the effects of a policy such as charter schools and remedial training in one setting does not guarantee that the policy will have the same effects in another setting (Banerjee et al., 2007; Cohodes and Parham, 2021).

policies subsequently implement measures to reduce health disparities, for example by expanding access to certain treatments, results often differ from expectations because the treatments turn out to perform less well among the new subgroups (Essien et al., 2021; Degtiar and Rose, 2023). Likewise, in development economics, a policy-maker may want to implement a program to improve the nutritional status of pregnant women in her country. Yet, the only available evidence may come from other behavioral intervention programs that happened not to include pregnant women (Duflo et al., 2007).

Transportability is increasingly attracting the attention of researchers in many policy-related fields (Pearl and Bareinboim, 2011; Westreich et al., 2017; Dahabreh and Hernán, 2019; Degtiar and Rose, 2023). Narrowly speaking, it refers to the degree to which internally valid evidence on the effects of an intervention derived for a study population can be extended to infer its effects on a particular target population of political interest that were not part of the original study population. More broadly speaking, it refers to the degree to which results from a policy experiment in context A can be expected to hold in context B (Marchionni and Reijula, 2019; Francesconi and James, 2021). Beyond these narrow and broader conceptualizations of transportability lie the possible welfare effects ('do no harm') of not taking poor transportability into account when designing or changing policies.

Current calls for energy efficiency assistance programs targeted at low-income households also have to contend with problems of transportability. Most evidence on the effectiveness of policies to increase household energy efficiency derives either from observational data or RCTs from untargeted programs in which low-income households are absent or, at a minimum, systematically under-represented. Household income appears to be a relevant dimension for transportability of behavioral insight: Studies in other areas, for example health insurance (Domurat et al., 2021) and consumer debt (Holzmeister et al., 2022), have found that low-income households respond differently to seemingly well-established behavioral insights, sometimes even in the opposite direction. Deploying behavioral insights to the energy efficiency investment choice of low-income households could therefore also give rise to null effects or even negative outcomes.

In this paper, we empirically examine the transportability of behavioral insights in the context of an initiative aiming to improve one of the world's largest energy efficiency assistance schemes. One program within this scheme, the Refrigerator Replacement Program (RRP), offers cash incentives to low-income households for replacing old and inefficient refrigeration appliances with new, highly efficient ones. Improvements to the RRP were sought against a backdrop of modest program performance as seen by management and sponsor: A take-up rate of about 25 percent among all eligible households was regarded as improvable given that program eligibility requires passing an individualized cost-benefit test and that verifying eligibility is a major cost factor for the program. The performance of the RRP has proven to be responsive to small changes in its design: In the past, unsystematic procedural changes had substantial impact on program take-up (Chlond et al., 2022), opening a window for targeted experimentation. The RRP's management agreed to partnering with researchers in order to estimate treatment effects of several candidate program improvements and compare those with the average treatment effect of management's own baseline, rather than bringing in an outside consultancy to receive a single "behaviorally informed" improvement proposal. The possible shape and form of candidate improvements was limited, however, by stringent design constraints common to the public sector, such as adherence to administrative procedures and not adding expenses to program administration (Della Valle and Bertoldi, 2021).

The paper reports on the resulting study, which takes the shape of a co-designed RCT that compares the effects of eight treatments. Six treatments are candidate improvements to be tested for transportability, one is the management's own program baseline, and one is a legacy design. The candidate improvements had two targets: One target was the presentation of the appliance replacement opportunity in the "program information letter", a critical feature in many public programs (Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019; Hotard et al., 2019; Linos et al., 2022). This letter is handed to households on the occasion of an audit visit and is intended to make them enrol in the RRP based on a solid understanding of the program. A review of the literature identified three main variations that had a track record of positive treatment effects on households' energy savings elsewhere[2] while being consistent with the design constraints. One was a visual enhancement of the economic effects of the replacement opportunity (Allcott and Greenstone, 2017; Stojanovski et al., 2020). In part, this variation was already included in the management's own baseline. The second variation was the introduction of loss framing (Gonzales et al., 1988; Homar and Cvelbar, 2021) in how the replacement opportunity is presented. The third variation was the use of peer experience with the appliance replacement (Allcott, 2011; Andor et al., 2020) when presenting the RRP in lieu of an individualized forward projection. Notwithstanding their general track record, these variations are essentially untested in the context of low-income households because such households are typically under-represented in this literature (Allcott, 2011; Fowlie et al., 2015).

The other target of candidates for improving the RRP was post-visit engagement with the low-income households. Here, a review of literature identified the use of reminders through letters, text messaging, and visual cues left in the household as a likely program improvement. This was based on amassed evidence on the positive effects of reminders on program take-up (Guyton et al., 2016; Gravert, 2021) and a literature that shows that low-income households are particularly likely to be subject to salience problems and cognitive scarcity (e.g., Shah et al. 2012; Haushofer and Fehr 2014), therefore potentially benefiting from repeated engagement some time after receiving the information letter. As before, however, the studies documenting the effectiveness of reminders contained few to no observations from low-income households.

Our experimental design does not only allow for an examination of the transportability of tried

[2]See Khanna et al. (2021) for a recent meta analysis.

and trusted candidate improvements to the RRP, but also generates effect estimates that themselves make progress against established criteria of external validity, such as the 'SANS' (selection, attrition, naturalness and scalability) conditions (List, 2020). We randomly selected 21 trial sites for participation in our natural field experiment (Harrison and List, 2004). All treatment variations had to survive a demanding co-design process involving program management and program staff to ensure effortless administration, high naturalness, and full scalability to all 150 program sites in the country, including a pilot trial at one site. Throughout the entire experimental process, we continually raised awareness among program and site managers to disclose potential attrition. No attrition was reported to us among treated units.

Based on the observed behavior of 1,803 low-income households over the course of one year, we conclude that the transportability problem is both real and consequential when trying to design energy efficiency assistance programs targeted at low-income households. The unique opportunity afforded to us by program management succeeds in testing – with respect to take-up – candidate program improvements that are consistent with the program constraints, are scalable, and have a high degree of external validity. That test shows, however, that none of the candidate improvements outperforms management's own baseline, a visually enhanced presentation of the replacement opportunity without a reminder. Worse, some candidate improvements significantly decrease program take-up, with unexpectedly bad performance for loss framing the replacement opportunity. Unexpectedly, adding reminders can also backfire. The evidence points to the most vulnerable households among the low-income households as those program participants for whom transportability breaks down to the point of reversing the direction of treatment effects. The evidence on framing is consistent with the idea that in terms of prospect theory, low-income households tend to use reference points for assessing gains and losses from investing in energy efficient appliances that are significantly shifted to the left of that used by the average household behavioral insights have been obtained from (Kőszegi and Rabin, 2006). It is also consistent with evidence that loss tolerance is more widespread than previously considered and more prevalent among those with experience of adverse financial shocks (Chapman et al., 2024). The evidence on reminders is consistent with the idea that the announcement of a reminder leads to lower task completion because anticipated reminders undermine own investment in imperfect memory (Ericson, 2017). This may be particularly relevant to cognitively stressed low-income households. Our results add important texture to the demands of "Do No Harm" policy making: Should it be the case that the most vulnerable groups are most at risk from poor transportability, then this poses additional challenges for policies supposed to address their needs because additional safeguards and tests could restrict options and cause delays.

The paper is structured as follows. Section 2 highlights our contribution to the existing literature. Section 3 describes our experimental design, providing detailed information on the procedures of the energy efficiency assistance scheme, the RRP and the subsidy voucher (Section 3.1), on the treatment development and rationale (Section 3.2), and on the selection and training of the program sites where we conducted the RCT (Section 3.3). We next describe our household sample in Section 4 and our empirical strategy in Section 5. The presentation of the results follows three steps. First, we focus on the effects of the treatments targeting the information letter in Section 6.1, and second, we focus on the effects of the reminder effects targeting the post-visit engagement in Section 6.2. Third, Section 6.3 focuses on heterogeneous treatment effects by the type of federal income support received, which we use as a proxy of vulnerability and aspiration differences among our low-income sample. Section 7 concludes.

2 Related literature

Our study lies at the intersection of three strands of literature. First, our paper contributes to the emerging literature investigating the effects of behavioral interventions on the take-up of governmental welfare and public assistance programs. Recent studies investigate take up of Earned Income Tax Credit (EITC) benefits (Linos et al., 2022; Bhargava and Manoli, 2015), unemployment benefits (Bruckmeier et al., 2021), the SNAP food stamp program (Finkelstein and Notowidigdo, 2019), claims for tax refunds (Bronchetti et al., 2013), waivers for citizenship applications (Hotard et al., 2019) or medicare insurance (Brot-Goldberg et al., 2023). As a common finding of these evaluations, take-up rates are usually rather modest. For example, only 14 percent claim the EITC benefits (Bhargava and Manoli, 2015) and 6 percent of eligible low-income households enrol in the SNAP (Finkelstein and Notowidigdo, 2019). At the same time, this literature reports mixed evidence on the effectiveness of different behavioral interventions including presentational changes in program description in letters and reminders on program take-up among low-income households. Bhargava and Manoli (2015) observe that enhancing information letters by simplifying the explanation of process and salient program benefits can increase EITC take-up from 14 to 31 percent. In the context of fee waiver applications, Hotard et al. (2019) show information nudges to be effective as they increase application rates by 8.6 percentage points. Similarly, Finkelstein and Notowidigdo (2019), find positive effects of an information letter on SNAP take-up, in particular if the information letter further includes assistance information. In contrast, Linos et al. (2022) report null-effects of information letters that vary content, design, messenger and mode on EITC take-up. Some studies additionally examine the role of default sets: While Brot-Goldberg et al. (2023) show that default rules can have large and persistent effects on enrollment and drug utilization in a voluntary drug benefits program, Bronchetti et al. (2013) find no significant default effect among low-income tax filers. We contribute to this literature in several ways: One is the setting, which brings international evidence from Germany to bear on the question of how take-up can be improved. Another is the nature of the assistance program: The RRP does not provide support in cash or in kind. Instead, it incentivizes investment decisions with medium-term cash-flow benefits to the household. Such programs raise new issues that merit attention due to their potential for sustained improvements in household finances, but also due to non-trivial welfare aspects of diverting cash from consumption to investment.

Second, the paper contributes to the literature on transportability of empirical evidence across study settings, in particular in public policy (Pearl and Bareinboim, 2011; Dahabreh and Hernán, 2019; Degtiar and Rose, 2023). One manifestation of a transportability problem is when priors for expected effect sizes that derive from existing empirical evidence are subsequently confronted with divergent evidence from a new study population. The inconsistent results reported by the literature on take-up of assistance programs discussed above are a case in point that transportability cannot be assumed. Like generalizability, transportability is an aspect of the external validity of extending inferences beyond the study sample, but is distinct from the former (Westreich et al., 2017): The problem of generalizability arises when the study sample is a strict and possibly non-random subset of the target population. The problem of transportability arises when the study sample is not a subset of the target population (Duflo et al., 2007).[3] When there is insufficient overlap, an internally valid sample average treatment effect (SATE) of a policy may not allow valid inference to its specific target average treatment effect (TATE) (Goldstein et al., 2019). Understanding more about the extent to which policies designed for particular target groups can rely on insights from interventions tested elsewhere is urgently required (Halpern and Sanders, 2016; Hallsworth, 2023). This is particularly relevant when the specific target group consists of vulnerable people, given the possible ethical implications of a failure to transport (Harrison et al., 2020). We contribute to this literature by conducting an RCT in which we explore the transportability to low-income households of behavioral insights that have consistently performed successfully in the general population. By showing that transporting some of these insights not only fails to improve performance, but can lower it, we demonstrate the importance of transportability in behavioral public policy.

The third strand of literature to which our experiment contributes deals with the impact of behavioral interventions on energy savings and efficiency, but with a particular focus on low-income households. While information nudges perform well in the field of energy efficiency investments within broader population samples (Khanna et al., 2021), little is known on the effectiveness of behavioral interventions on investments in low-income households. One particular reason is that empirical evidence from non-targeted energy savings programs provide only limited insights as low-income households are more likely to drop out of those programs compared to an average household. For example, Löschel et al. (2023) find that low-income households are less likely to adopt a cost-free energy saving app. Allcott (2011) reports that low-income households are more likely to stop receiving the cost-free home energy reports (HERs), which aligns with follow-up work stressing a relatively lower willingness to pay for such reports among tenants (Allcott and

[3]While agreeing on the lack of overlap between study and target population (Hotz et al., 2005; Allcott, 2015), the literature has not yet converged on a single definition of transportability. See Dahabreh and Hernán (2019) and Degtiar and Rose (2023) for a discussion.

Kessler, 2019). Closest to our work are studies by Fowlie et al. (2015, 2018) reporting low take-up of financial incentives among low-income US households for energy efficient building weatherization, even though the gains of doing so are high. Using a randomized encouragement design Fowlie et al. (2015) find that despite massive additional expenditures ($1,000 per audited household), audit take-up only moderately increases from 1 percent in the control to 6 percent in the encouraged group. The energy efficiency literature has recently started to provide a more nuanced picture on the transportability of behavioral interventions to new study populations. For the case of HERs, Andor et al. (2020) report the effects of social comparison-based HERs on residential electricity consumption to be significantly lower for targets groups other than US residents. Similarly, Bonan et al. (2021) show that prime-augmented HERs may even backfire if they address customers who hardly engage in pro-environmental behavior. Our study contributes by providing additional evidence from low-income households on the heterogeneity of treatment effects in the context of energy efficiency. This evidence not only includes the presence of weak, but also the presence of negative effects on program take-up.

3 Experimental Design

We implement our RCT within the largest energy efficiency assistance scheme in Germany, the "Energy-Saving-Check." In the following sections, we first describe the program (3.1) and then turn to the experimental variations and the hypotheses (3.2). After this, we explain the roll-out of the experiment in the selected local program sites (3.3).

3.1 The "Energy-Saving-Check" and the "Refrigerator Replacement Program"

The "Energy-Saving-Check" (SSC, German: *Stromspar-Check*) is a nation-wide program that aims at lowering the energy bills of low-income households in Germany by reducing their electricity and water consumption. The SSC is implemented jointly by the German Caritas Association, one of the largest social welfare organizations in the country, and the Association of Energy and Climate Protection Agencies (eaD). Annual funding of around 10-15 million Euro is provided by the German Federal Ministry for the Environment on the basis of program grants with a funding cycle of three years, subject to successful (re-)application by the implementing agencies. Within the SSC, the Refrigerator Replacement Program (RRP; German: *Kühlgeräte - Tauschprogramm*) has been offering cash vouchers to households on federal income support[4] in order to encourage replacing their old and inefficient refrigeration devices with modern, highly efficient models. The RRP started on January 1, 2009 and was scaled up to its current size with the start of the second funding cycle of the SSC ("SSC plus") in April 2013.

The recruitment of qualified households into the SSC's home energy audit takes place through a variety of channels. The program is actively promoted in employment and social assistance agencies through both printed and audiovisual material. In addition, active staffers provide individualized descriptions of the program using pop-up booths in shopping streets and malls. Some local branches of the social assistance agency mandate the participation of households with excessively high energy bills. The SSC also maintains a website where information is available about the RRP in eleven languages. Additionally, recruitment takes place directly through the local branches.

The typical home energy audit of the SSC consists of two visits to the household by a two-person team within a period of around three weeks. During the first visit, the "energy advisors" make an inventory of all electric devices and their usage in the household, assess the electricity consumption of refrigerators and freezers, and educate the household on electricity-saving behavior. The inventory and electricity consumption assessment are used to screen for eligibility of the household for the RRP. The screening leads to differences in the second visit: Both eligible and non-eligible households receive approximately 70 Euros worth of energy-saving kit such as LED light bulbs, switchable socket strips, TV standby cut-off switches, timers and water flow regulators. These items are directly installed by the two advisors. Non-eligible households then exit the SSC initiative. For eligible households, the second visit contains an additional component in which they are specifically targeted for enrolment in the RRP through educational material and promotion.[5]

The rationale for enrolling households in the RRP is the large contribution, roughly 25 percent (BDEW, 2019), that refrigerators make to the electricity consumption of the average German household.[6] Differences in refrigerator efficiency can therefore significantly impact residential electricity bills. To be eligible for enrolment, the low-income household has to own a refrigerator older than 10 years and be expected to save at least 200 kWh annually from a replacement with the most energy efficient class of devices on the market.[7] The expected financial savings are communicated to

[4]To qualify, the household needs to receive federal income support such as unemployment benefits ("Arbeitslosengeld II"), housing allowances ("Wohngeld," "Sozialhilfe"), low pensions ("Grundsicherung"), child supplements ("Kinderzuschlag") or benefits for asylum seekers ("Leistungen nach Asylbewerberleistungsgesetz"), or the household's income must be below the income limit for attachment. In 2020, more than 7 percent of German households qualified on this basis (Bundesagentur für Arbeit, 2020).

[5]Only households that completed the first visit of the SSC home energy audit can become eligible for the RRP.

[6]We use "refrigerator" to refer to both refrigerators, freezers, and combination units within the program.

[7]The savings expectations are based on engineering estimates: Based on the inventory data from the first visit, SSC staff use a custom database to calculate expected savings based on a comparison between the current device and a reference device of equivalent size and features that fulfills the A+++ standard, the most efficient class of devices on the EU scale in force between 2009 to 2021. Since March 2021, a revised EU scale has been in force that puts devices previously rated as A+++ in classes A to F. Transitional arrangements were in place both in the retail sector and in the RRP.

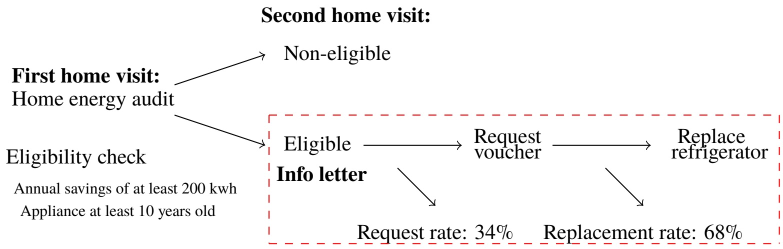


Figure 1: Procedure of the home audits.

the household via an information letter during the second visit. The investment of *every* eligible household passes a cost-benefit evaluation. Figure 1 summarizes the procedure of the two visits.

Figure 2 displays the letter that all eligible households receive during the second visit. It contains all relevant information and provides the basis for our experimental variations. First, it informs households that they meet the necessary eligibility criteria for the RRP and provides an estimate of the expected annual electricity savings (in EUR) from successful replacement (see green box in Figure 2). Second, it offers a step-by-step explanation on how to request and redeem the voucher: (1) eligible households *request* the voucher at one of the local sites, (2) they *replace* their refrigerator with a new model and (3) *redeem* the voucher in cash after successful replacement. To be able to redeem the voucher, a number of criteria have to be met: Households need to present their purchase receipt, document that the purchased device is of EU Energy Label class A, B, C or D; and provide proof that the old refrigerator has entered the recycling chain. Households have to handle all steps of the refrigerator replacement on their own, including identifying and selecting a model that fulfils the requirements, (pre-)financing the purchase, and organizing the logistics of delivering the new and of disposing of the old refrigerator. Once requested, the voucher is valid for two months. Consequently, there exists a sharp 'deadline' on when the voucher expires.

Our experimental interventions vary the presentation of key information, as – common in many public policy assistance programs – the letter is a crucial bottleneck in the process. As Figure 1 shows, only 34 percent of all eligible households request the voucher. Conditional on the request, 68 percent of the households then successfully replace their refrigerator. The overall replacement rate of 23 percent consequently substantially falls short of the 100 percent replacement target that the program designers optimally would aim for considering the cost-benefit evaluation of the investment that every eligible household passed.

Recipient
Street name, House number
Postcode City
Telephone number
Stromspar-Check

strom spar-check.de

Mr / Ms / Family
Surname
Street name, House number
Postcode City

Important information from Stromspar-Check Aktiv on exchanging your refrigerator

Dear Mr/Ms/Family ...,

We measured your old refrigerator as part of the Stromspar-Check ('energy efficiency check') and, when analysing the values, found out that it would be worth exchanging your appliance.

If you purchase a new, highly efficient appliance, you could save around XX euros per year in electricity costs!

If you exchange your old appliance, you will receive a subsidy of 100 euros* for the purchase of your new appliance. We will need to issue you a voucher for this amount <u>before</u> you buy the new appliance. The number of vouchers is limited. Please contact your nearest Stromspar-Check location before you purchase.

It is <u>extremely important</u> that you note the following points:

- Your new refrigerator <u>must have</u> the energy efficiency class **A, B, C or D** of the new energy efficiency label (valid from March 1, 2021) and may not exceed energy consumption of **xxx kWh** per year.
- As soon as you have found a suitable appliance and you can finance it, call your nearest location (telephone number: xx/xxxxx). The Stromspar team will issue an individual voucher and send it to you. Please note that the **voucher is only valid for two months from the date of issue (up to a maximum of 11.02.2022).** It is <u>not</u> possible to extend this period if the voucher has expired!
- **Only buy your new refrigerator <u>after</u> you have received the voucher from us, because the purchase date has to be within the voucher's validity period.**

Please also note:

- Your old appliance has to be properly disposed of and we require written proof of disposal.

After purchase and disposal, you can redeem your voucher from the nearest Stromspar-Check location. To do so, you need to bring the following:

- the voucher
- the original proof (or receipt) of proper disposal
- the original purchase receipt of the new refrigerator (we will make a copy)
- the energy efficiency label of the new refrigerator (A, B, C or D sticker for the energy efficiency class)
- your ID card or passport.

Should you have any questions or require further information, feel free to contact us at your nearest location before your purchase.

Best regards,
Your Stromspar-Check Team

* = We cannot pay more than the purchase price of the refrigerator. The voucher does not cover, as applicable, ensuing disposal, transport and delivery costs.

Figure 2: Letter informing about voucher eligibility.

Note: Green highlights are added to the original letter. Red highlights are not added and part of the original letter.

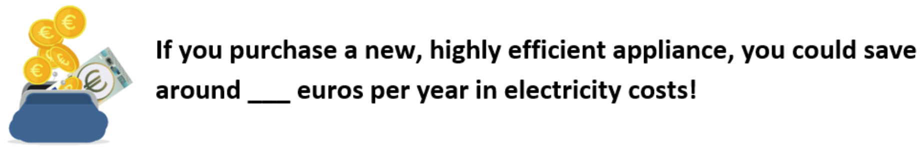
3.2 Treatments and Hypotheses

In a co-design process with the program officials and local site managers we jointly developed and pre-screened a set of treatment variations within or alongside the information letter. The aim of this procedure was to ensure effortless administration, high naturalness, and full scalability to all program sites in the country. Consequently, in April 2020, we started with a set of interactive online workshops where we jointly identified potential barriers for assessing the most economically relevant information provided in the letter and discussed a first set of potential interventions. It turned out that sensible language is a very important condition.[8] The treatments were then refined in a series of follow-up workshops and pre-tested in a pilot on site (in Frankfurt) from July to December 2021. Based on the lessons-learned from the pilot, we organized a final workshop in March 2022 to present the finalized experimental design and the procedures to the sites selected to participate in the RCT (see Section 3.3). The roll-out started in April 2022.

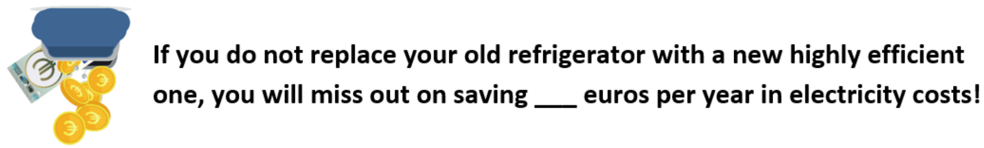
Inspired by well-established insights from behavioral economics, our interventions target three main dimensions. First, we distinguish whether we frame the reported annual savings in electricity cost as a financial GAIN or as a financial LOSS if households miss the opportunity of replacement. Here we differentiate between our first three treatments. The GAIN treatment simply reflects a legacy version of the information letter as shown in Figure 2. They key sentence in GAIN reads "If you purchase a new, highly efficient appliance, you could save around ... euros per year in electricity costs!" (green box). The RRP management planned to extend this legacy condition by adding a small purse pictogram with money falling in to the left of the text (see Figure 3a) and to introduce this extended version as the new baseline, replacing the legacy version. We refer to the management's own baseline version as GAIN^{+} treatment. In contrast to the management's baseline, a LOSS^{+} treatment points out the expected foregone savings in annual electricity costs from non-replacement, with the purse pictogram rotated by 180 degrees and money falling out (see Figure 3b). Here, the key sentence reads as follows: "If you do not replace your old refrigerator with a new highly efficient one, you will miss out on saving ... euros per year in electricity costs!"

As a second dimension, we randomize whether the reported annual savings in electricity costs stem from individual-level, appliance-specific engineering estimates (as in the legacy and baseline versions of the information letter) or from actual replacements being recently conducted in households with similar characteristics (i.e., with respect to the composition of household members). The basic idea of the first set-up is to calculate the expected annual savings in individual electricity costs based on a comparison between the current actual electricity consumption levels of the old device and a hypothetical reference device of equivalent size and features that meets the necessary efficiency levels. For realizing these financial gains, it is assumed that households exactly follow

[8] E.g., we proposed a social norm intervention in the spirit of Allcott (2011), which was rejected by site managers. They were afraid of social pressure resulting from the letter and an additional burden placed on the low-income households.



(a) GAIN$^+$ Frame.



(b) LOSS$^+$ Frame.

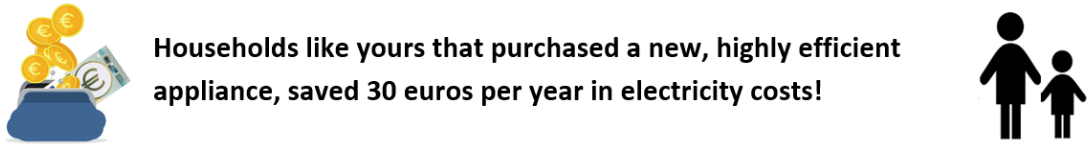
Figure 3: Treatment variations: Gain and Loss Frame.

the suggested protocol and purchase a model similar to the reference device provided, that they use it in an optimal sense and that there is no change in individual electricity prices. In contrast, in the second set-up, our ‘peer experience’ treatments display electricity savings based on realized monetary values based on actual replacements by a peer group with a similar household composition. We calculate the annual savings realized by determining the average difference in energy costs between the old appliance and the new appliance actually purchased for the following six household types, representing 89 percent of all households in our sample: single person (26 EUR savings); two adults (29 EUR savings); two adults and one child (30 EUR savings); two adults and two children (36 EUR savings); two adults and three or more children (41 EUR savings); single parenting (30 EUR savings).

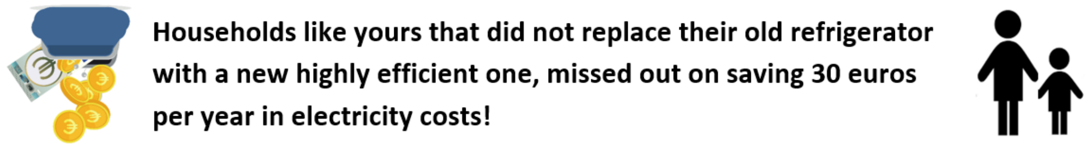
We highlight the reference to peer-behavior in two ways. First, we add a pictogram of the respective household composition to the right side of the savings information to the letter. Second, we alter the text to read “Households like yours that purchased a new, highly efficient appliance, saved ...” (GAIN$^+$ PEER) as opposed to “If you purchase a new, highly efficient appliance, you could save ...” (GAIN$^+$) (see Figure 4a). The alterations are similar in the loss frames (see Figure 4b).

As a third treatment dimension, we introduce different reminders for eligible households (see Figure 5). In a first variant, in accordance with the EU General Data Protection Regulation (GDPR), households are asked for their written consent to be recontacted 4-8 weeks after having received the information letter. Conditional on consent[9] and depending on households’ stated preferences, local site managers then send out a letter- or SMS-based reminder at the beginning of a new month (when households usually are more financially liquid; see Figure 5a for a translation of the reminder

[9]To our knowledge, every household provided consent.



(a) GAIN$^+$ PEER Frame.



(b) LOSS$^+$ PEER Frame.

Figure 4: Treatment variations: Peer experience Gain and Loss Frame.

text). In a second variant, the energy advisor places a tag, which displays the logo of the Energy-Saving-Check program, inside the refrigerator during the second visit (see Figure 5b).[10]

We organize these three treatment dimensions in a "2x3 + Reminder" design, as displayed in Figure 6. The columns of the table capture the first treatment dimension, the variation in the framing. We distinguish between (1) the legacy GAIN frame, (2) the visually enhanced management's baseline frame, i.e., GAIN$^+$, and (3) the visually enhanced LOSS$^+$ frame. The rows of the table in Figure 6 display the second treatment dimension, i.e., the savings based on individual estimates or peer experience. While the legacy treatment is only combined with the individual savings estimate, we vary for both the GAIN$^+$ and the LOSS$^+$ frame whether they are combined with the individual savings estimate or the peer-experienced savings. Finally, we combine selected treatments with the third dimension, the reminders. That is, the orange fields in Figure 6 display treatment versions that we test both with and without reminders.[11]

In sum, we implemented a total of eight different treatments: (1) GAIN (legacy version), (2) GAIN$^+$ (management baseline), (3) LOSS$^+$, (4) GAIN$^+$ PEER, (5) LOSS$^+$ PEER, (6) GAIN$^+$ REMINDER, (7) LOSS$^+$ REMINDER, (8) LOSS$^+$ PEER REMINDER. Due to technical issues with the database, the legacy GAIN treatment could only be implemented starting at the end of July 2022. To meet our scalability targets, the randomization process was carried out automatically by

[10] During the co-design process, we discussed whether the tag should be placed inside or outside the refrigerator. The program officials and site managers insisted on placing the tag inside the refrigerator as an outside-placement may give rise to stigma. Persons visiting the audited household would directly see the logo of the program directed to low-income households.

[11] Note, that this definition is agnostic with respect to the specific reminder types. Such differentiation between reminder types is only relevant for the GAIN$^+$ treatment, which we combined with (i) the SMS/letter reminder, (ii) the refrigerator tag and (iii) both reminder types. In Figure 6 and the regression analysis, we pool these three groups into one GAIN$^+$ REMINDER treatment. By contrast, both reminder treatments in the loss domain of Figure 6 are combined with the refrigerator tag, but abbreviated in our specifications as LOSS$^+$ REMINDER and LOSS$^+$ PEER REMINDER treatment, respectively. In robustness checks, we differentiate between the reminder versions.

Reminder for the refrigerator exchange in Stromspar-Check Aktiv

Dear Mrs./Mr./Family xxx,

we would like to remind you that you can request a voucher of 100€ for the purchase of a refrigerator of energy efficiency label class A, B, C or D as part of the Stromspar-Check. To do so, please contact your location.

With kind regards

Your Stromspar-Check Team

(a) SMS or letter reminder.

(b) Refrigerator tag.

Figure 5: Reminder treatments.

	Gain		Loss
Individual estimates	GAIN (legacy)	GAIN$^+$ (baseline)	LOSS$^+$
Peer experience		GAIN$^+$ PEER	LOSS$^+$ PEER

+

REMINDER

Figure 6: Treatment dimensions

the program database. Households are allocated to the different treatments with equal probabilities, except for (6) GAIN$^+$ REMINDER to which we over-sampled.[12]

With our experimental design, we test four main hypotheses. Our first hypothesis concerns the impact of using visual enhancement or visual aids in an information letter about energy efficiency in order to reinforce the verbal message (Stojanovski et al., 2020; Allcott and Greenstone, 2017). In our case, the visual enhancement consists of adding the purse pictogram to the letters, to emphasize the monetary consequences of program participation. We expect that such enhancements will be conducive to increasing attention to the message of the letter, in particular the reference to the euro-denominated savings in the message body.

H1: Refrigerator replacement rates will be higher in GAIN$^+$ compared to GAIN .

The second hypothesis concerns possible performance differences between the GAIN$^+$ frame and an alternative LOSS$^+$ frame. Such message framing, in particular the framing of the same material outcome as a gain or a loss by shifting the mental reference point, has received attention in the literature on enhancing household energy efficiency for some time. A starting point is the seminal study by Gonzales et al. (1988) who examined the effect of exposing 408 home owners who qualify for enrolment in a energy efficiency retrofit program to one of two different Home Energy Audit procedures. In one, home owners were visited by auditors trained to employ a gain framing by referring to the benefits of enrolling; in the other, by auditors trained to employ a loss framing by referring to the foregone benefits of not enrolling. They find significantly higher enrolment among home owners in the loss framing treatment. This finding seems to reflect a higher impact of loss framing on behavior in general (see Kühberger (1998) for a meta-study) and in the specific context

[12] The reason is that the GAIN$^+$ REMINDER treatment was originally designed to feature two treatments, one using an SMS reminder, the other a mailed letter reminder. Feedback during the pilot phase led to the decision to combine the two treatments into one, with the household choosing how the reminder would be provided. In addition, as discussed above, treatment group (6) includes participants who are reminded by the SMS/letter, participants who are reminded by the tag and participants who are reminded by both the tag and the SMS/letter.

of energy savings (see Homar and Cvelbar (2021) for a meta-study covering 61 studies).[13]

The cumulative evidence that favors loss over gain framing in the context of energy efficient behavior leads us to hypothesize that the visually enhanced loss frame will give rise to a higher probability that households will replace their refrigerators than the visually enhanced gain frame. Both treatments increase the salience of estimated savings via the added pictogram, but in line with much of the literature, we expect the loss frame to provoke stronger behavioral reactions.

H2: Refrigerator replacement rates will be higher in the LOSS⁺ frame compared to the GAIN⁺ frame.

One threat to H2 comes from the lack of specific evidence on how low-income households respond to different message framings. This is relevant because most of the cumulative evidence on message framing is derived on the basis of the average household, with little discussion on its "transportability", i.e., on whether the same patterns also apply to our specific low-income demographic. Their behavioral patterns could differ.[14]

Thirdly, we hypothesize a reference to peer experiences to both reduce uncertainty on the actual electricity savings from replacement and to increase the personal relevance of the program. First, when expected electricity savings are calculated based on individual projected estimates, it may not be entirely clear to the household what "*can* save you" (as indicated in the information letter) exactly means in this context. By contrast, the peer experience design is more specific in that it provides estimates of *realized* savings by similar households. Second, considering the large literature on peer effects in energy behavior (e.g., Allcott 2011, Andor et al. 2020), households may have higher trust in information stemming from households in similar socioeconomic circumstances.[15] Hence, the peer experience may decrease the degree of uncertainty and increase the attachment households place on the possible savings.

H3: Refrigerator replacement rates will be higher in the peer experience design compared to a design relying on individual projected estimates.

The fourth hypothesis focuses on the effect of reminders as a tool to overcome potential procrastination of, and inattention to, the replacement choice. There is previous evidence from the RRP that inattention and procrastination may have a role to play in explaining low levels of replacement

[13] Recent studies by DeGolia et al. (2019); Park et al. (2023) point in the same direction. The finding is not universal, however: Some studies also find no (e.g., Sussman et al., 2018) or the opposite (e.g., Chen, 2023) effect.

[14] For example, Mullainathan and Shafir (2013) show scepticism towards the effectiveness of loss frames among low-income individuals. Relatedly, Fehr et al. (2022) show that financial scarcity decreases the likelihood to exhibit an endowment effect, a mainstay of behavioral biases among experimental subjects.

[15] Likewise, also research in other domains provides evidence of social learning from peer behavior, e.g., Escobar and Pedraza (2023) in the context of stock trading and Abdulai (2023) in the context of agricultural farming technologies.

among eligible households. Specifically, Chlond et al. (2022) find that inadvertent program changes in the RRP that also involved having to set households deadlines for replacing their refrigerators led to increased take-up. This is a pattern that would be consistent with households suffering from procrastination, which can be overcome by deadlines as a form of goal setting.

Reminders are a popular strategy for overcoming procrastination and inattention among targeted individuals. This popularity can be explained by the effectiveness of reminders across a wide range of behavioral contexts (see Gravert (2021) for a recent review), with even small reminders frequently resulting in sizeable effects. This also holds for the context of energy efficiency (Fang et al., 2023) and of low-income households (Karlan et al., 2016; Guyton et al., 2016). The behavioral economics of reminders emphasizes limited attention (Karlan et al., 2016) and the interplay of present bias and limited memory (Ericson, 2017) as drivers of why households pay insufficient attention to the future benefits of an action (here, the future saving from replacing the appliance), overemphasize the current cost (here, the outlay for replacing the appliance), leading them to postpone – and ultimately forego – an otherwise beneficial investment. Reminders intervene in this process by lowering the cost of attention and/or overcoming limited memory. On this basis, we predict that the effect of reminders on replacement behavior is positive.

H4: Refrigerator replacement rates will be higher in the reminder treatments compared to the non-reminder treatments.

Threats to H4 come from two different areas. One area is the fact that in our experiments, reminders are always combined with a gain or a loss framing. The reminder effect may therefore depend on the initial frame presented in the information letter, and thereby on the first-order effects of this framing (see H2). At least two issues arise as a result. One, the frame could affect the composition of the sample being recontacted for the reminder.[16] Two, the frame could determine the direction of the reminder effect, reinforcing or possibly reversing the effect.[17]

The other area of threat is the procedural requirement of compliance with the EU GDPR in programs such as the RRP, ruling out collecting and using contact details for re-contacting household without legitimate cause and prior informed consent. Prior informed consent to being reminded could lead to an 'anticipation' effect that could exacerbate rather than mitigate procrastination (Ericson, 2017). The reason is that anticipating future reminders, the household will find it in its interest to allocate even less costly mental effort to acting on the decision situation than when not anticipating being reminded. This anticipation effect may indeed be negative and contribute to

[16]For example, if H2 holds and low-income households have higher replacement rates under a loss than under a gain frame, then the sample that has not redeemed the voucher yet and, hence, receives the reminder will not be the same across the two framings.

[17]For example, if H2 does not hold and households are discouraged through a loss frame, then a household can make sure that it will not be exposed to a loss frame again by replacing the refrigerator. Therefore, the added reminder may reverse the framing effect and increase the incentives for replacement.

lower replacement rates in those reminder treatments that ask for permission to receiving a letter or SMS message. The treatment that uses a tag, on the other hand, would be expected to be unaffected from such anticipation effect since the tag is present from the beginning and for as long as the household leaves it in place.

3.3 Selection and Training of Sites

To address potential concerns of site selection and to support the scaling idea of our interventions, we randomly selected a number of local sites being invited to participate in our field experiment. To this end, we collected data on the number of long-term unemployment benefit recipients of the 150 municipalities in which an SSC field office is located. We determined weights that reflect the share of the number of benefit recipients in the service area of a site relative to the total number of benefit recipients in Germany. Higher weights were consequently assigned to sites which cover a greater number of benefit recipients. This weighting ensures that each individual benefit recipient has the same probability of being part of our study. The selected sites were then determined by a weighted randomized draw of a pre-arranged number of 30 out of 150 total sites.

The randomly drawn sites received invitations by the program officials to participate in our study in December 2021. 23 out of the 30 sites followed our invitation.[18] One site (Munich) had to be excluded from participation due to a large-scale investment program tested in parallel there. A second site (Groß-Gerau-Kreis) experienced an unexpected lack of site management, which is why this site was as well excluded from the study. Three additional sites (Jena, Weimar, Erfurt) as well as the pilot site (Frankfurt) showed interest and self-selected themselves into the experiment. To avoid resulting selection effects, we exclude these sites from our analysis.[19] Hence, our final sample consists of the following 21 sites: Anklam, six sites in Berlin, Bremen, Gelsenkirchen, Hamburg, Ibbenbüren, Cologne, Konstanz, Leipzig, Meißen, Mettmann (county), Minden, Offenburg, Osnabrück (county), Saarlouis, Wiesbaden. Figure 7 displays all program sites (pink), the sites participating in our study (red) and the excluded self-selected sites (violet).

After the selection process had been determined, we organized various training sessions to familiarize the site managers with the treatments and their implementation. This was necessary as all treatments, except for the legacy GAIN treatment, were new to the site managers participating in the RCT. The managers of the selected sites are distinct from the site managers participating in the co-design process, and were not informed about the origins and discussions underlying the different treatments (including management's request for the baseline GAIN$^+$ frame). As a further important aspect of scalability, the program database was adapted to assign eligible households to the different treatments and automatically print out the correct version of the information letter. Thus, there is no leeway for site manager choices.

[18] The sites not signing up were Chemnitz, Delmenhorst, Essen, Recklinghausen, Moers and two sites in Dortmund.
[19] Results are however robust to including them.

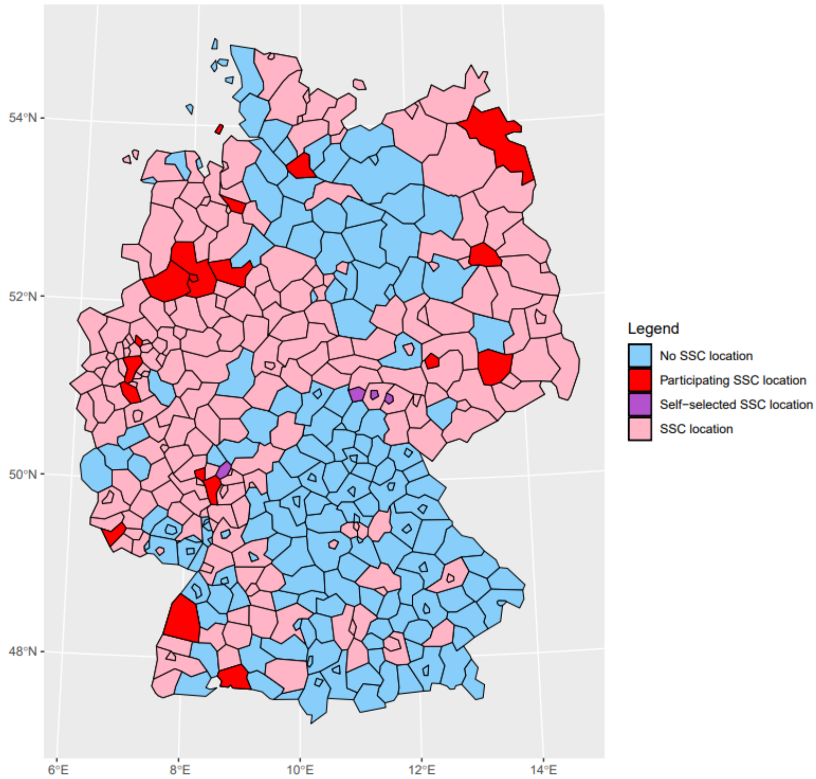


Figure 7: Program sites of the "Energy Saving Check".

However, the site managers are responsible for sending the SMS and letter reminders, and for equipping the auditors with the refrigerator tag. For this reason, we prepared an online tutorial, Q&A sessions and information material for the site managers. The Q&A sessions were held from December 2021 to January 2022. From January 2022, the database programming was finished to include all randomized treatments and all study sites received treatment materials. The period between January 2022 and April 2022 thus served as an additional trial phase for site managers to inform, organize and train their local staff to the interventions.

4 Sample Description

Our study sample consists of program participants, who (i) received their audit during the intervention period from April 1, 2022 to February 15, 2023,[20] (ii) were audited within one of our randomly selected intervention sites, and (iii) were found eligible for the refrigerator replacement program. In total, we observe 1,803 households that fulfill these criteria.

Table 1 displays summary statistics of household information being recorded in the program database after the first household visit. The average electricity price in the billing period 2021/2022

[20] The program's funding period ended on March 31, 2023, but the local sites stopped voucher issuance on February 15, 2023, due to accounting procedures.

Table 1: Summary statistics of household data

	Mean	Std. Dev.	Min	p10	p25	p50	p75	p90	Max	N
Electricity price (Euro/kWh)	0.33	0.05	0	0.30	0.30	0.32	0.34	0.40	1	1,803
Energy consumption (kWh)	2926.10	1708.17	0	1250.00	1712.00	2572.33	3700.00	4957.83	17,501	1,762
Electric water heating: (1=yes)	0.37	0.48	0	0.00	0.00	0.00	1.00	1.00	1	1,803
Annual est. savings (kWh)	336.33	153.37	12	204.67	228.68	293.64	414.62	547.56	1,466	1,802
No. persons (count)	2.78	1.88	1	1.00	1.00	2.00	4.00	5.00	11	1,803
Living space (m^2)	71.08	27.60	20	43.00	52.80	65.00	83.00	107.00	300	1,803
Unemployment benefits (1=yes)	0.65	0.48	0	0.00	0.00	1.00	1.00	1.00	1	1,803

Note: Displayed are summary statistics for the variables electricity price (in Euros/kWh), household energy consumption (in kWh), an indicator for heating warm water with electricity, engineering estimates of savings from refrigerator replacement (in kWh), number of persons in household, living space, and an indicator for receiving long-term unemployment benefits.

was 33 cents/kWh, which aligns with the average electricity prices paid nation-wide (Federal Statistical Office of Germany, 2023b). Electricity consumption in our sample is notably lower compared to national statistics. While the participants of the Energy-Saving-Check consume on average 2,926 kWh per year, the German average is at 3,383 kWh per year (Federal Statistical Office of Germany, 2023c). An important determinant of electricity consumption is whether warm water is produced using electricity which applies to 37 percent in the sample. The annual energy savings from refrigerator replacement as estimated by engineering estimates are on average 336.33 kWh. Together with the average electricity price, this maps into annual financial savings of 112.74 Euros on average. The average number of household members in our sample is 2.8 persons – slightly higher than the German average of 2.0 household members (Federal Statistical Office of Germany, 2023a). Despite larger household sizes, the living space of our sample (71.1m^2 on average) is lower compared to the national average (96.2m^2, Federal Statistical Office of Germany (2023d)).

These differences to national statistics are highly plausible given that all of our study participants are recipients of federal income support. The majority of 65 percent are recipients of long-term unemployment benefits (Arbeitslosengeld II), the second largest fraction of 16 percent receive a basic pension (Grundsicherung) and 12 percent receive housing benefits (Wohngeld).

We next turn to comparing the summary statistics across treatment groups. Table 2 displays the mean difference in characteristics between the respective treatment and (the mean of) all other groups. Due to randomization, in expectation we should see no major differences in observable characteristics between treatments. Importantly, we compare the individual estimated energy savings in kWh, as measured during the first home visit, and not the financial savings as communicated to the households. We would thus not expect differences in the estimated kWh savings across treatment groups.

The number of observations per treatment is about equally split across groups except for two exemptions. First, the number of observations for the GAIN group is lower due to technical issues causing a delay in implementation (see Section 3.2). Data collection for this group only started by the end of July 2022. Second, the number of observations is higher for the GAIN$^+$ REMINDER

Table 2: Differences in summary statistics by treatment group

Variable	GAIN	GAIN$^+$	LOSS$^+$	GAIN$^+$ PEER	LOSS$^+$ PEER	GAIN$^+$ REMINDER	LOSS$^+$ REMINDER	LOSS$^+$ PEER REMINDER
Energy price (ct/kWh)	0.017***	-0.008*	-0.004	0.008**	-0.005	-0.001	-0.007	0.007
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Energy consumption (kWh)	-52.260	219.960*	-60.091	-126.032	-84.370	77.374	-92.195	-79.227
	(169.694)	(132.055)	(142.873)	(137.835)	(145.405)	(83.596)	(143.276)	(144.543)
Electric water heating: (1=yes)	0.049	-0.022	0.025	0.056	-0.021	-0.030	-0.032	0.046
	(0.048)	(0.037)	(0.040)	(0.038)	(0.041)	(0.023)	(0.040)	(0.041)
Estimated savings (kWh)	37.361**	-0.413	-0.397	0.310	16.051	-18.097**	1.631	10.618
	(15.134)	(11.767)	(12.742)	(12.143)	(12.884)	(7.401)	(12.815)	(12.925)
No. persons (count)	-0.157	0.151	-0.001	-0.101	0.127	-0.009	0.017	-0.072
	(0.186)	(0.144)	(0.156)	(0.149)	(0.158)	(0.091)	(0.157)	(0.158)
Living space (m^2)	-2.150	2.861	0.768	-3.001	-3.551	1.723	-0.925	-0.009
	(2.727)	(2.116)	(2.292)	(2.184)	(2.318)	(1.333)	(2.306)	(2.326)
Transfer scheme (1-7)	0.382	-0.177	0.182	-0.129	-0.244	0.036	0.132	-0.101
	(0.314)	(0.244)	(0.264)	(0.252)	(0.267)	(0.154)	(0.266)	(0.268)
N per group	109	190	159	177	155	702	157	154
N total	1,803	1,803	1,803	1,803	1,803	1,803	1,803	1,803

Note: Displayed are the differences in means for the variables energy price, energy consumption, an indicator for heating warm water with electricity, estimated savings in kWh from refrigerator replacement, number of persons in household, living space, an indicator for receiving unemployment benefits and transfer type categories by treatment group compared to all other groups. Standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

group, since we summarize three initially implemented treatment groups that vary in the reminder version and we oversampled the GAIN$^+$ SMS/Letter REMINDER group as explained in Section 3.2.

Comparing the means of the respective covariates across groups, we see slight differences in the electricity price paid and the estimated kWh savings from refrigerator replacement. As one example, subjects in the GAIN treatment on average pay 0.17 ct more per kWh and their expected savings from replacement are on average 37.36 kWh higher compared to the average household in all other treatments (Table 1, column 1). These differences become insignificant or only marginally significant once adjusting for multiple hypothesis testing (List et al., 2019). We also add controls for these variables in our robustness checks.

Figure 8 displays the mean and 95 percent confidence intervals of our main outcome, refrigerator replacement, by treatment group. Replacement rates, i.e. the share of households who were informed about the subsidized replacement opportunity during the second visit and actually replaced their refrigerator, range between 14 and 24 percent. We observe the highest replacement rate among households randomly placed in the baseline treatment GAIN$^+$. The LOSS$^+$ treatment has the lowest replace rate. Between these two endpoints are the treatments using peer experience to encourage take-up, repeating the pattern of gain versus loss framing, as well as the legacy design GAIN. As Figure 8b shows, replacement rates do not substantially improve by combining changes in the framing of the information letter with reminders. In fact, reminders could negatively affect take-up: The replacement rate among households in the GAIN$^+$ REMINDER treatment is 17 percent, below the replacement rate of 24 percent in the GAIN$^+$ treatment. The heterogeneity of replacement rates across treatments is already visible at the voucher request stage, which must precede replacement. The voucher request rates range between 26 and 37 percent.[21] The request rate is highest in the GAIN$^+$ treatment and lowest in the LOSS$^+$ treatment.

In numbers, we observe that 46 out of 190 possible refrigerators are actually replaced in the GAIN$^+$ treatment. In the Loss treatment, only 22 out of 159 possible refrigerators are replaced. A first statistical comparison already indicates that the replacement rate of the GAIN$^+$ treatment is significantly higher compared to the average replacement rate of all other treatments (two-sided t-test, $p = 0.0203$). We investigate these surprising patterns closer in the next sections.

5 Empirical Strategy

To analyze the effect of the different behavioral interventions on RRP performance, we compare our two outcomes (voucher request and refrigerator replacement) across the treatment groups. That is, we first assign an indicator variable for each treatment group and then run a regression of household (i) voucher request and (ii) refrigerator replacement choice on the treatment indicators as follows:

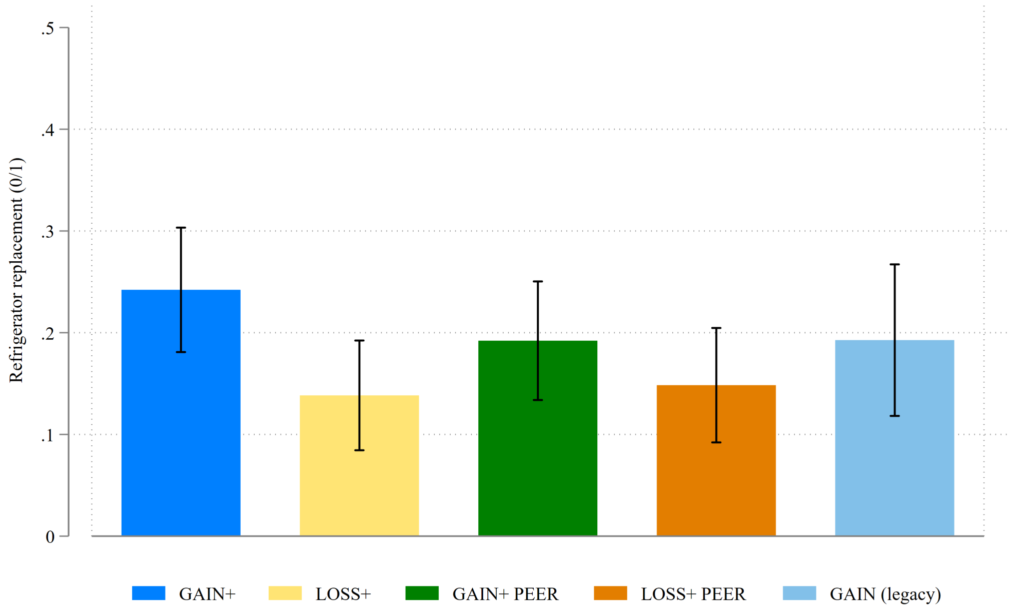
$$\begin{aligned} Y_i = & \beta_0 + \beta_1 GAIN_i + \beta_2 LOSS_i^{+} + \beta_3 GAIN^{+}PEER_i + \beta_4 LOSS^{+}PEER_i \\ & + \beta_5 GAIN^{+}REMINDER_i + \beta_6 LOSS^{+}REMINDER_i + \beta_7 LOSS^{+}PEER\ REMINDER_i \\ & + Savings\ Info_i + \mathbf{X}_i + \mathbf{F}_i + \varepsilon_i, \end{aligned} \tag{1}$$

whereby i denotes the individual household and Y_i refers to voucher request or refrigerator replacement choice. The treatment indicators equal 1 if the household is in the respective treatment group and 0 otherwise. We estimate equation (1) as a linear probability model (LPM).

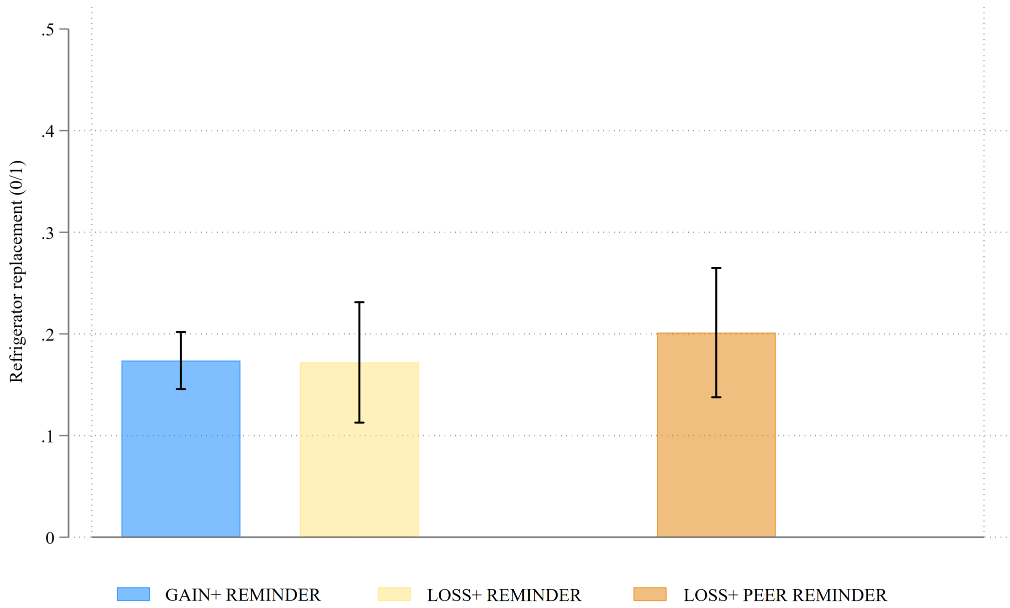
We define the GAIN$^+$ treatment as the omitted treatment category as the program management designated this treatment to be the new program baseline. Moreover, using the GAIN$^+$ treatment as baseline provides greater statistical power compared to using legacy program baseline, the GAIN treatment. Due to the delay in implementing the GAIN treatment, the group has the fewest observations. Using the GAIN treatment as baseline would thus penalize the power of all comparisons. We pool the different reminder versions that are combined with the GAIN$^+$-frame (GAIN$^+$ REMINDER) to increase power for analysis. However, in an additional robustness check, we explore potential differences between the tag and the SMS/Letter reminders.

In all regressions, we control for the expected financial savings from replacement as communicated to the household, denoted by *Savings Info*. Importantly, this variable adjusts for differences in communicated savings that might otherwise bias the comparison between the individual estimates and peer experience treatments. In controlling for *Savings Info*, we allow the different framing to impact replacement choices, but hold constant the displayed monetary values.

[21] Figure A.1 displays the corresponding voucher request rates.



(a) Information letter treatments.



(b) Reminder treatments.

Figure 8: Refrigerator replacement rates by treatment group.

Further, in subsequent specifications, we add control variables obtained from the program database. Specifically, the vector $\mathbf{X}_i$ summarizes the household's electricity price, electricity consumption, usage of electric water heating, the number of persons in the household, the living space, the type of social benefit transfer the household receives and the federal state the household lives in. Further, the vector $\mathbf{F}_i$ summarizes fixed effects for the local site, the two energy advisors auditing the household, month fixed effects and month-site fixed effects.[22] Finally, ε_i denotes the error term.

6 Results

We first discuss the results on the treatments that modify the information letter design in Section 6.1. Results on the reminder treatments are discussed in Section 6.2. Section 6.3 investigates heterogeneous treatment effect by the type of federal income support received.

6.1 Information letter treatments

Table 3 displays the regression results for voucher request across the different treatment groups and Table 4 shows the corresponding results for refrigerator replacement. In both tables, the GAIN⁺ treatment is the omitted baseline group.

Table 3 shows that both the legacy GAIN and the LOSS⁺ frame lead to lower voucher request rates compared to the management's baseline GAIN⁺ treatment. In column (1), these reductions are significant at the 10- and 5-percent level, respectively. Further, across the different models the lower request rate in the LOSS⁺ treatment remains robust and significant at the 5-percent level. The impact of the frame on behavior is substantial: Framing the reported annual savings in electricity costs as a financial loss hampers the conversion rate of an information letter into a voucher by 10.5 to 11.5 percentage points. The exception is column (4), in which the LOSS⁺ coefficient reduces to 5.4 percentage points and turns insignificant. Column (4) includes advisor fixed effects, and thus controls for the specific pair of advisors that visited the household. However, once including the advisor fixed effects, we loose variation in our outcome variable, which likely explains the loss of significance of some treatments. We thus view specification (5) as our preferred specification. It includes control variables as well as site, month and month-site fixed effects, but disregards the advisor fixed effects. Figure 9a graphically summarizes the estimates reported in column (5).

For the PEER-treatments, we find mainly positive coefficients, suggesting larger voucher request rates compared to the GAIN⁺ treatment. Yet, these coefficients are not significant. Specifically, for the GAIN⁺ PEER treatment, the coefficients in column (1)-(4) suggest 0.2 to 5.6 percentage points higher request rates but are not statistically significantly different from zero. In our preferred

[22]We include month-site fixed effects since some sites have complementary programs that increase the voucher value and vary over time. E.g., the Berlin sites introduced additional cash vouchers in November 2022. These complementary programs exist in at least four of the sixteen states and in a number of municipalities.

Table 3: Treatment effects on voucher request

	(1)	(2)	(3)	(4)	(5)
	Voucher request (0/1)				
$GAIN^{+}$	REF	REF	REF	REF	REF
$LOSS^{+}$	-0.105**	-0.115**	-0.108**	-0.054	-0.111**
	(0.050)	(0.049)	(0.047)	(0.050)	(0.047)
$GAIN^{+}$ PEER	0.031	0.056	0.002	0.025	-0.003
	(0.053)	(0.054)	(0.055)	(0.058)	(0.057)
$LOSS^{+}$ PEER	-0.004	0.032	0.004	-0.017	0.014
	(0.054)	(0.055)	(0.053)	(0.055)	(0.053)
GAIN (legacy)	-0.098*	-0.085	-0.077	-0.040	-0.071
	(0.056)	(0.055)	(0.054)	(0.054)	(0.053)
$GAIN^{+}$ REMINDER	-0.065*	-0.062	-0.068*	-0.038	-0.059
	(0.039)	(0.039)	(0.037)	(0.037)	(0.036)
$LOSS^{+}$ REMINDER	-0.044	-0.048	-0.044	-0.029	-0.041
	(0.051)	(0.050)	(0.047)	(0.047)	(0.047)
$LOSS^{+}$ PEER REMINDER	0.031	0.071	-0.032	-0.017	-0.034
	(0.055)	(0.056)	(0.055)	(0.057)	(0.056)
Constant	0.301***	0.541***	0.309***	0.324***	0.415***
	(0.043)	(0.085)	(0.045)	(0.045)	(0.100)
Savings Info	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No	Yes
Fixed Effects	No	No	Yes	Yes	Yes
Advisor FE	No	No	No	Yes	No
N	1802	1761	1785	1725	1745

Note: Linear probability models of voucher request (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The $GAIN^{+}$ treatment is the omitted reference treatment group. All regressions control for the communicated savings from replacement. Columns (2) and (5) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state, the social benefit transfer scheme, and whether the household heats warm water with electricity. Columns (3)-(5) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Column (4) controls for the two advisors visiting the household. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

Table 4: Treatment effects on refrigerator replacement

	(1)	(2)	(3)	(4)	(5)
	Refrigerator replacement (0/1)				
$GAIN^{+}$	REF	REF	REF	REF	REF
$LOSS^{+}$	-0.104**	-0.117***	-0.102**	-0.062	-0.113***
	(0.042)	(0.041)	(0.040)	(0.044)	(0.040)
$GAIN^{+}$ PEER	-0.044	-0.045	-0.061	-0.037	-0.074
	(0.046)	(0.046)	(0.046)	(0.049)	(0.046)
$LOSS^{+}$ PEER	-0.088*	-0.065	-0.071	-0.065	-0.056
	(0.045)	(0.045)	(0.046)	(0.048)	(0.046)
GAIN (legacy)	-0.051	-0.051	-0.046	-0.025	-0.046
	(0.049)	(0.048)	(0.048)	(0.052)	(0.047)
$GAIN^{+}$ REMINDER	-0.069**	-0.072**	-0.075**	-0.051	-0.070**
	(0.034)	(0.034)	(0.032)	(0.035)	(0.032)
$LOSS^{+}$ REMINDER	-0.070	-0.078*	-0.082**	-0.066	-0.085**
	(0.043)	(0.044)	(0.042)	(0.044)	(0.042)
$LOSS^{+}$ PEER REMINDER	-0.035	-0.020	-0.090*	-0.070	-0.098**
	(0.047)	(0.048)	(0.046)	(0.049)	(0.047)
Constant	0.234***	0.292***	0.233***	0.226***	0.238**
	(0.037)	(0.072)	(0.039)	(0.040)	(0.096)
Savings Info	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No	Yes
Fixed Effects	No	No	Yes	Yes	Yes
Advisor FE	No	No	No	Yes	No
N	1802	1761	1785	1725	1745

Note: Linear probability models of refrigerator replacement (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The $GAIN^{+}$ treatment is the omitted reference treatment group. All regressions control for the communicated savings from replacement. Columns (2) and (5) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state, the social benefit transfer scheme, and whether the household heats warm water with electricity. Columns (3)-(5) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Column (4) controls for the two advisors visiting the household. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

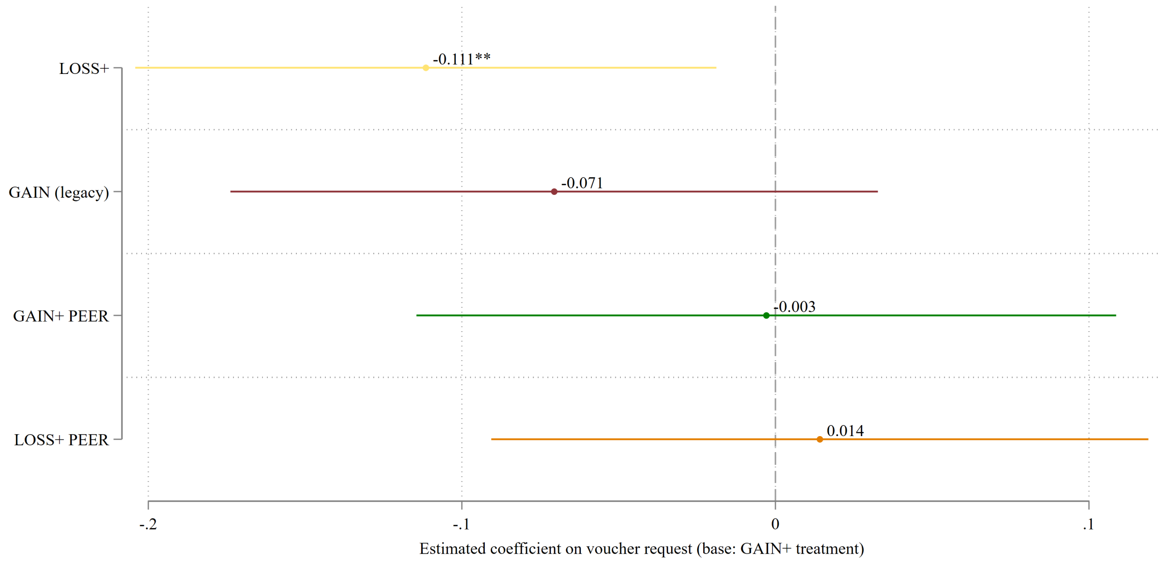
specification (5), the coefficient even turns negative but remains small and indistinguishable from zero. The coefficient of the LOSS+ PEER treatment in specification (5) suggests a 1.4 percentage points higher request rate compared to the GAIN+ treatment, but is insignificant from zero. In the other specifications, the LOSS+ PEER coefficients range from –0.017 to 0.032 and are, yet again, not statistically different from zero.

As displayed in Table 4, for our main outcome of interest, refrigerator replacement, all estimated treatment coefficients display a negative sign and, thus, indicate that the GAIN+ baseline yields the highest refrigerator replacement rates. Again, the difference to the LOSS+ treatment is most robust, yielding a statistical significant difference at the 1- or 5-percent level, indicating that replacement rates are 10.2-11.7 percentage points lower as compared to the GAIN+ treatment. As discussed above, the only exception is the specification in column (4). When adding the advisor fixed effects, the negative effect of LOSS+ reduces to 6.2 percentage points and becomes insignificant. We suspect a substantial decrease in the variation of replacement rates across treatments once conditioning on advisor pairs as explanation for this significance loss. We thus view column (5) as our preferred specification. Figure 9b visualizes the coefficients reported in column (5) in Table 4.

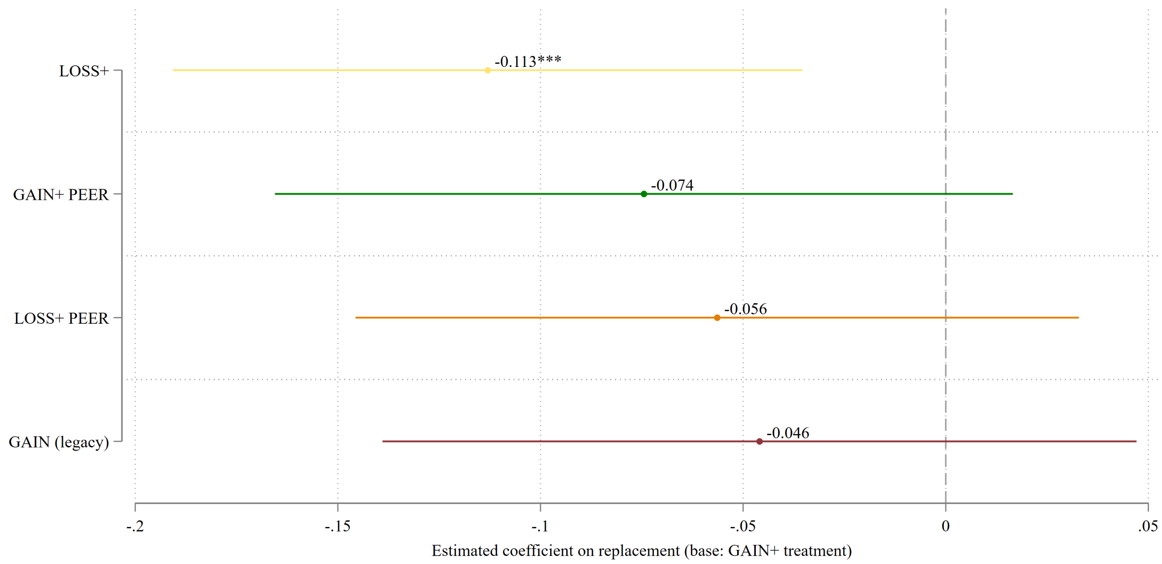
With respect to our hypotheses, we do not find robust evidence supporting H1. While replacement rates are 4.6-5.1 percentage points higher in the management's newly introduced GAIN+ frame compared to the legacy GAIN frame, this difference is not statistically significant. Moreover, our findings clearly oppose our second hypothesis H2. The LOSS+ frame significantly reduces replacement compared to a GAIN+ frame. Here, our findings thus contrast with evidence on the effectiveness of loss frames (Laibson and List, 2015), but support recent literature raising doubt in the universal belief in loss frames (Gal and Rucker, 2018).

The comparison of Table 3 and Table 4 reveals a nuanced picture of providing information of peer experiences: While the mainly positive coefficients of the PEER treatments suggest an increase of voucher request rates relative to the GAIN+ baseline, the peer experience seems to decrease refrigerator replacement rates. The coefficients of both the GAIN+ PEER and the LOSS+ PEER treatments are negative in Table 4. For the LOSS+ PEER treatment, the negative coefficient is even significant at the 10-percent level in the specification (1) of Table 4.

Table A.1 explores this linkage and displays the results of a regression of the replacement probabilities conditional on requesting the voucher. The coefficients of the GAIN+ PEER and the LOSS+ PEER treatments are significant, large and negative. This implies that although the peer experience tends to encourage households to request the voucher – potentially because they are more optimistic regarding their own replacement choice – these households finally fail to successfully conduct the replacement. In other words, peer experience information may increase the rate of requested vouchers, but it does not lead to more energy efficient appliances. From a program perspective, such peer information effect may be rather undesirable: More vouchers are in circulation, which increases the administrative effort and liabilities on the program's balance sheet, but no replacements follow.



(a) Voucher request.



(b) Refrigerator replacement.

Note: Displayed are the estimated coefficients of column (5) of Table 3 and column (5) of Table 4. The coefficients are sorted by effect size.

Figure 9: Coefficient plot of information letter treatments.

In summary, the results displayed in Table 4 fail to confirm H3. The peer experience information in the GAIN+ PEER treatment does not increase replacements as compared to the individual engineering estimates in the GAIN+ baseline. The coefficients rather point to a negative effect but do not statistically differ from zero. Likewise, post-estimation tests both on column (1) and (5) fail to reject equality of the LOSS+ and LOSS+ PEER treatments ($p = 0.7111$ and $p = 0.1784$, respectively). Hence, despite the size of coefficients points towards the LOSS+ PEER treatment realizing higher replacement rates compared to the LOSS+ treatment, which would be confirmatory of H3, we do not see robust statistical evidence.

Overall, we find the GAIN+ treatment to be the most effective information letter design. It tends to improve upon the simpler GAIN version and substantially improves upon the LOSS+ version. Further, information about successful peer experience cannot foster technology adoption in our sample.

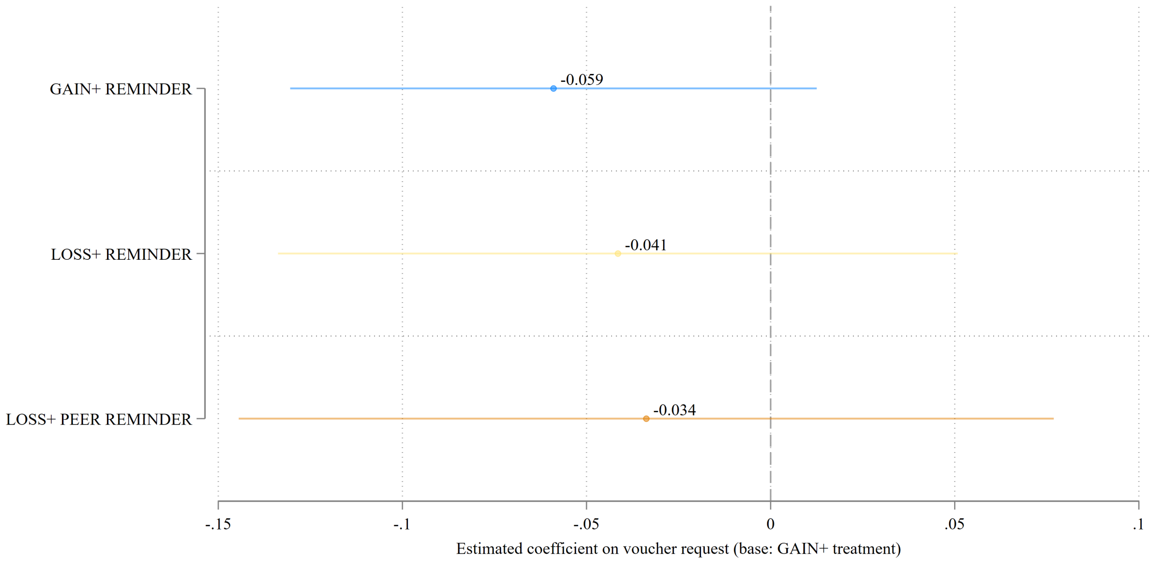
6.2 Reminder treatments

Figure 10 displays the estimated coefficients for the reminder treatments of column (5) of Table 3 in the upper panel and Table 4 in the lower panel. The estimated effects on voucher request are negative but mostly insignificant. More interestingly, for refrigerator replacement, we observe a significant and negative effect for the GAIN+ REMINDER treatment. At the 5-percent significance level, the added reminder decreases energy efficiency investments by 6.9-7.5 percentage points compared to the GAIN+ baseline in all specifications, except for the previously discussed model (4).

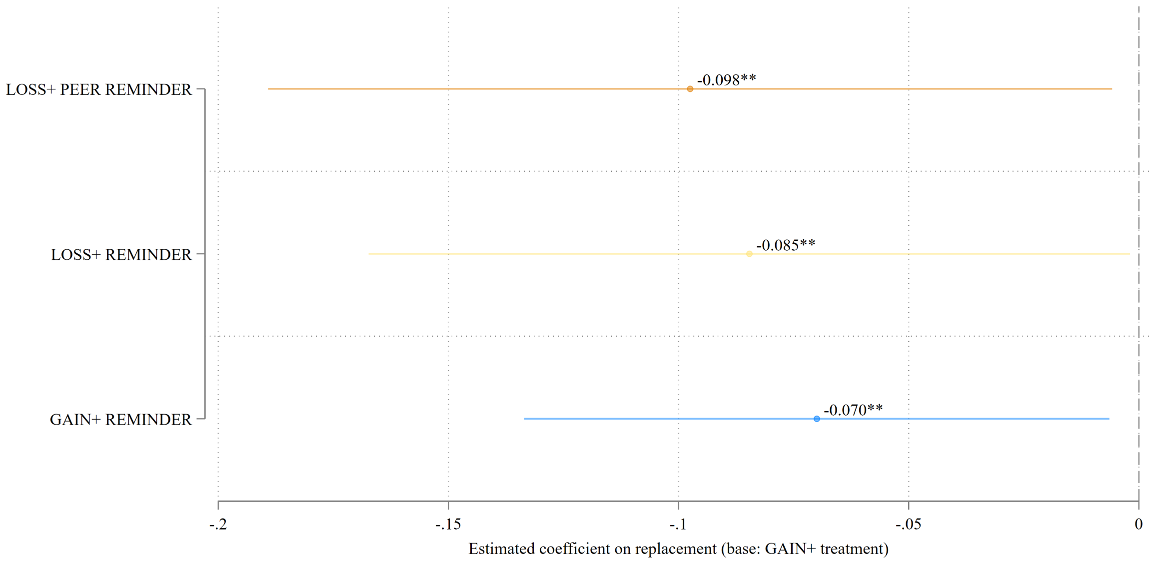
This finding is in stark contrast to our hypothesis H4. The reminder decreases the effectiveness of the GAIN+ treatment. A potential explanation for such an empirical finding might be found in the negative anticipation effect as described in Section 3.2. The GAIN+ participants may have anticipated receiving a reminder on the replacement choice, thus postponed the decision, allocated less mental effort to remember, and, ultimately, never followed through with the investment choice.

At first sight, this finding is similar for the combination of the loss framing and the reminder. As indicated in Figure 10 and in our preferred specification (5), both the LOSS+ REMINDER and the LOSS+ PEER REMINDER coefficient are negative and significant at the 5-percent level. Across our different specifications, the effect of the LOSS+ REMINDER treatment ranges between a 7.0 and 8.5 percentage points decrease in replacement rates compared to the GAIN+ baseline. The LOSS+ PEER REMINDER treatment causes 2.0-9.8 percentage points lower replacement rates. However, the significance of these effects varies and depends on the inclusion of fixed effects.

Further, to evaluate the LOSS+ REMINDER and the LOSS+ PEER REMINDER coefficients with respect to H4, we need to compare them to the LOSS+ and LOSS+ PEER treatments. According to a post-estimation test on column (5) of Table 4, the LOSS+ REMINDER treatment does not significantly improve upon the LOSS+ treatment ($p = 0.4833$), and the LOSS+ PEER REMINDER



(a) Voucher request.



(b) Refrigerator replacement.

Note: Displayed are the estimated coefficients of column (5) of Table 3 and column (5) of Table 4. The coefficients are sorted by effect size.

Figure 10: Coefficient plot of reminder treatments.

treatment does not significantly improve upon the LOSS+ PEER treatment ($p = 0.3404$).[23]

Hence, we likewise reject H4 for the loss treatments but, contrary to the gain treatments, do not find evidence consistent with negative anticipation effects. Importantly, one explanation for the significantly negative coefficient of the GAIN+ REMINDER treatment and the lack of significant difference between the LOSS+ REMINDER, the LOSS+ PEER REMINDER and their respective non-reminder treatment versions may be the use of different reminder formats. The LOSS+ REMINDER and the LOSS+ PEER REMINDER treatments introduce the fridge tag. This is different for the GAIN+ REMINDER treatment. This treatment group pools three reminder treatments, (1) a GAIN+ Letter/SMS REMINDER, (2) a GAIN+ Tag REMINDER and (3) a GAIN+ Letter/SMS and Tag REMINDER.

Table A.3 displays the replacement regression results for the different reminder versions.[24] In theses regressions, the GAIN+ Letter/SMS REMINDER and the GAIN+ Letter/SMS Tag REMINDER treatments robustly show significant and negative coefficients. The GAIN+ Letter/SMS REMINDER treatment effect ranges between a 7.2-8.2 percentage point decrease in the likelihood to replace the refrigerator, which is significant at the 5-percent level. For the GAIN+ Letter/SMS Tag REMINDER treatment, the coefficient displays a 8.8-10.4 percentage points decrease at the 1- to 5-percent significance level. By contrast, the GAIN+ Tag REMINDER treatment does not significantly reduce replacement rates compared to the GAIN+ baseline. Similarly, as discussed above, the LOSS+ Tag and LOSS+ PEER Tag REMINDER treatments do not significantly reduce replacement rates compared to the LOSS+ treatment and the LOSS+ PEER treatment, respectively. Hence, only the Letter/SMS reminders significantly reduce replacement rates and cause the rejection of H4. The tag reminders do not significantly affect replacement rates in comparison to their respective information letter-only version.

This finding is consistent with the anticipation effects proposed by Ericson (2017). Only the Letter/SMS reminder was announced to households, by contrast, the tag reminder was directly placed into the household fridge during the second home visit. Thus, only the Letter/SMS reminder might have caused households to anticipate the reminder, postpone the replacement choice and reduce the mental capacity spent on thinking about the replacement choice. Ultimately, this explains the backfiring effect that we observe in our data.

In the next section, we more closely investigate the role of our sample in an exploratory analysis. In particular, we explore the extent to which some of the treatment effects can be attributed to the specific decision-making processes of low-income households.

[23]The respective p-values of post-estimation tests on column (1) of Table 4 are $p = 0.4094$ for LOSS+ REMINDER vs. LOSS+ and $p = 0.2205$ for LOSS+ PEER REMINDER vs. LOSS+ PEER.

[24]Table A.2 displays the corresponding results for voucher request.

6.3 Heterogeneous Treatment Effects

We delve deeper into understanding the role that the vulnerable situation of households in our sample plays in explaining households' (non)reaction to the tested interventions by investigating heterogeneous treatment effects by federal income support type.[25] As described in Section 4, 65 percent of our sample receive long-term unemployment benefits (Arbeitslosengeld II, Bürgergeld since January 2023). These unemployment benefits are set to provide a minimum subsistence level only,[26] and, hence, involve the lowest benefit payments we observe in our sample jointly with low pension supplements and benefits for asylum seekers. Households that receive a housing allowance or child supplements earn a small own income that is not sufficient to finance their rent and heating cost and additional expenses for children, respectively.

Sociologists, for example, argue that households receiving long-term unemployment benefits tend to differ in their mindset and attitudes towards society from other groups of welfare recipients. This could lead to a different behavioral response to interventions. In this literature, the difference in mindset and attitudes is traced back to heterogeneity among recipients with respect to the legitimacy of receiving public benefits, as perceived by others and by themselves. These perceptions have turned increasingly negative, in particular for long-term recipients (Dörre, 2015), leading to feelings of discrimination and of a lack of solidarity towards the long-term unemployed (Köster, 2023). This leads to a mindset of feeling left behind, denied access to respectable segments of society, and pitted against other welfare recipients (Dörre, 2015).

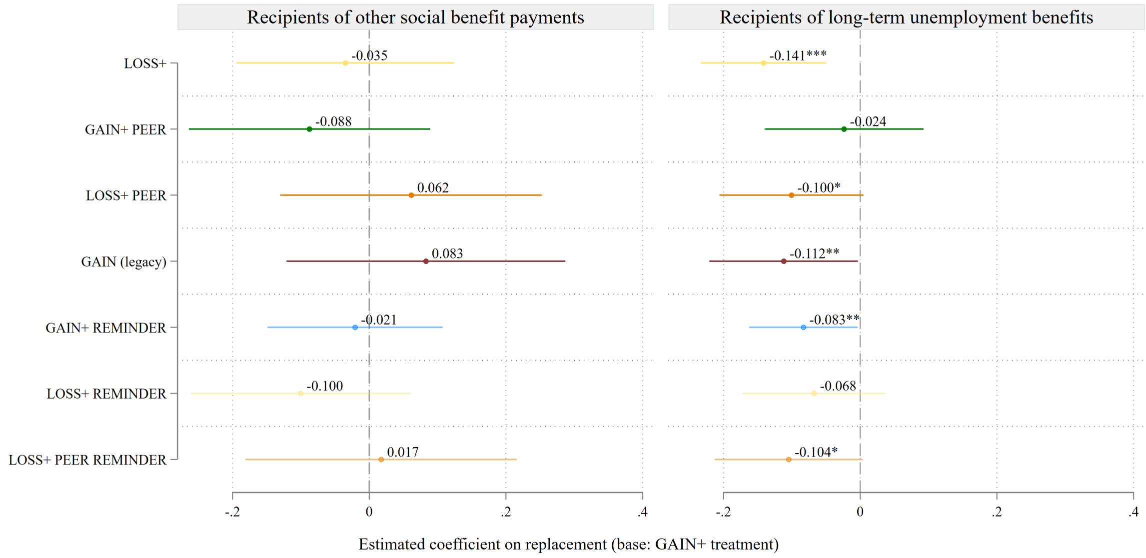
We compare the effectiveness of the treatments on refrigerator replacement rates between ALGII recipients and the recipients of other social benefit payments. To this end, we run regression (1) either among the sample of ALGII-recipients or among the recipients of other social benefit payments. Table A.4 displays the corresponding regression results. Figure 11 presents the estimated coefficients of column (4) for the sample of recipients of other social benefit payments and of column (8) for the sample of recipients of long-term unemployment benefits (ALGII).

Figure 11 shows evidence of heterogeneous treatment effects by transfer type. We observe insignificant effects of the treatments for the recipients of other benefit payments than the unemployment benefits on the refrigerator replacement decision.

By contrast, for the recipients of long-term unemployment benefits (ALGII), the GAIN$^+$ baseline significantly outperforms the LOSS$^+$ treatment – a finding that we observed on average but seems to be driven by the program participants who live off long-term unemployment benefits. The LOSS$^+$ coefficient is significant at the 1-percent level, and shows a 14.1 percentage points lower likelihood to replace the refrigerator compared to the management's baseline.

A similar finding holds for the GAIN$^+$ REMINDER treatment. The negative coefficient compared to the GAIN$^+$ treatment, which Table 4 displayed on average, is only significant for the sample

[25]Please note that this analysis was not pre-registered and is exploratory.
[26]In 2022, the long-term unemployment benefits are set to 449 Euros per month for a single-person household.



Note: Displayed are the estimated coefficients of column (4) (left panel) and column (8) (right panel) of Table A.4.

Figure 11: Coefficient plot of heterogeneous treatment effects on refrigerator replacement rates by transfer type.

of ALGII-recipients. More specifically, among that sample, the GAIN$^+$ treatment outperforms the refrigerator replacement rates of the GAIN$^+$ REMINDER treatment at 5-percent significance and suggests 8.3 percentage points higher replacement rates.

Further, in this analysis, the new GAIN$^+$ baseline also significantly outperforms the legacy GAIN treatment. Among the recipients of long-term unemployment benefits, GAIN decreases the likelihood of replacement by 11.2 percentage points, which is significant at the 5-percent level. For this particular sample, we can thus confirm the hypothesis H1.

Overall, the analysis by transfer types reinforces our conclusions from the average treatment effect analysis. The GAIN$^+$ baseline yields investment rates that are significantly higher than for the LOSS$^+$ treatment, particularly for the marginalized group of the long-term unemployed. Prior evidence and prospect theory would have predicted higher replacement rates in the LOSS$^+$ treatment. A potential explanation is that our specific target group interprets the frames differently from the average household because of their experience and their resulting beliefs in their ability to complete an appliance replacement. In this context, the GAIN$^+$ frame entertains a notion of the household succeeding in their planned action, in contrast with much of these households' lived experience, positively affecting beliefs. The LOSS$^+$ frame, on the other hand, reinforces the households' beliefs in their likely failure to complete the replacement. These changes in beliefs and expectations could explain why the loss framing backfires for the target group, in particular for long-term unemployed households. For the latter, the GAIN and the GAIN$^+$ REMINDER treatment also perform significantly worse that the GAIN+ baseline while there is no significant difference for other households. Both the positive effect of a visual enhancement as well as the backfiring of pre-announced reminders appears to be specific to this vulnerable group, highlighting its sociological specificity.

7 Conclusion

When trying to design or improve programs targeting specific groups in society, such as economically disadvantaged households, it is tempting for program managers to transport behavioral insights that have proven successful among the general population to their particular application context. This temptation is also present in the context of large-scale energy efficiency assistance programs that are regarded as underperforming, but whose performance has also been shown to respond favorably to small procedural changes. Yet, behavioral insights that have demonstrated their effectiveness for the general population may not generate the intended impacts among the target group, either in terms of magnitude or not even in terms of direction. When the target group is vulnerable or exposed, then the possibility that such transportation of insights has adverse impacts carries additional significance from an ethical perspective.

Partnering with the management to improve the Refrigerator Replacement Program, one of the world's largest energy efficiency assistance programs, we exploited the opportunity to test a set of alternative candidate improvements based on behavioral insights. These improvements needed to conform with the stringent requirements set by managers: Adherence to existing administrative procedures, a track record of effectiveness elsewhere, and zero or negligible cost. To be included among the alternatives were a baseline incorporating visual elements as well as a legacy design. In addition, we constrained ourselves to alternatives and procedures that allow us to fulfill emerging standards of scalability, the so-called SANS conditions. The process of co-designing and piloting resulted in six co-designed treatments, on top of the the baseline and the legacy design, that incorporated behavioral insights. Visual enhancement, loss framing, and using peer-based comparison were chosen to improve the information stage of the program while reminders were intended to improve the post-visit engagement of the target households. The evidence generated under conditions of randomized site selection, absence of attrition, demonstrated naturalness, and scalable interventions affirms that the transportability of behavioral insights to low-income households is under threat. While visual enhancements delivered largely as expected, peer experiences turned out to be ineffective as an improvement. Loss framing in fact back-fired in terms of program performance. Reminders were at best ineffective.

Our results can be seen to have implications for both researchers and policy-makers in the present, the short and the long run. In the present, they point to the need for researchers to disclose more actively the heterogeneity of treatment effects that certain behavioral interventions generate, to acknowledge the absence of evidence on treatment effects for specific groups, and to concede threats to the scalability of treatment effects derived without adherence to conditions such as SANS. Policy-makers for their part need to strive for a better understanding of the evidence that is used to derive average treatment effects and advocate for policy changes and for a better understanding of how this evidence relates to their specific target group.

In the short run, we believe that our results make the case for a more systematic collection of evidence on groups under-represented in studies of behavioral public policy in order to build up the 'true' distribution of treatment effects in the population as a whole. Researchers will need the cooperation from policy-makers to carry out such systematic collection. This requires an openness among policy makers to piloting and better yet to conducting controlled experiments yielding internally valid estimates of effects.

If undertaken, such joint efforts by researchers and policy-makers have a chance of resulting in a convergent evidence base that narrows confidence intervals around inferential estimates of transported policies. This chance relies on the premise that certain regularities are likely to arise, which will enhance the inference of effect sizes for target groups as more testing is conducted in the short run. The target groups of political import in many OECD countries tend to be vulnerable sections of society, in particular economically disadvantaged groups. These groups exhibit a high degree of heterogeneity across and within. At the same time, their members also confront similar cognitive and affective challenges. A widening evidence base among behavioral economists on the commonalities in how such members respond to different instruments will put researchers in a better position to predict and advise on promising strategies and for policy-makers to make more informed choices than at present.

References

Abdulai, A. (2023). Information acquisition and the adoption of improved crop varieties. *American Journal of Agricultural Economics 105*(4), 1049–1062.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics 95*(9-10), 1082–1095.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics 130*(3), 1117–1165.

Allcott, H. and M. Greenstone (2017). Measuring the welfare effects of residential energy efficiency programs. *NBER Working Paper 23386*.

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–276.

Andor, M. A., A. Gerster, J. Peters, and C. M. Schmidt (2020). Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management 103*, 102351.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Bhargava, S. and D. Manoli (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review 105*(11), 3489–3529.

Bonan, J., C. Cattaneo, G. d'Adda, and M. Tavoni (2021). Can social information programs be more effective? The role of environmental identity for energy conservation. *Journal of Environmental Economics and Management 108*, 102467.

Bronchetti, E. T., T. S. Dee, D. B. Huffman, and E. Magenheim (2013). When a nudge isn't enough: Defaults and saving among low-income tax filers. *National Tax Journal 66*(3), 609–634.

Brot-Goldberg, Z., T. Layton, B. Vabson, and A. Y. Wang (2023). The behavioral foundations of default effects: Theory and evidence from Medicare Part D. *American Economic Review 113*(10), 2718–2758.

Bruckmeier, K., R. T. Riphahn, and J. Wiemers (2021). Misreporting of program take-up in survey data and its consequences for measuring non-take-up: New evidence from linked administrative and survey data. *Empirical Economics 61*, 1567–1616.

Bundesagentur für Arbeit. Bedarfsgemeinschaften und deren Mitglieder 2020. https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html?nn=1460284&topic_f=gs-asu-sgbii-rev. Accessed: 2022-05-19.

Bundesverband der Energie- und Wasserwirtschaft e.V. (BDEW). Energiemarkt Deutschland 2019. https://www.bdew.de/service/publikationen/bdew-energiemarkt-deutschland-2019/. Accessed: 2022-05-19.

Chapman, J., E. Snowberg, S. Wang, and C. Camerer (2024). Looming large or seeming small? attitudes towards losses in a representative sample. *Review of Economic Studies*, rdae093.

Chen, C. (2023). Framing energy-efficiency programs: A survey experiment. *Energy Policy 183*, 113776.

Chlond, B., T. Goeschl, and M. Kesternich (2022). More money or better procedures? Evidence from an energy efficiency assistance program. *Alfred-Weber Institute for Economics, AWI Discussion Paper No. 716*.

Cohodes, S. R. and K. S. Parham (2021). Charter schools' effectiveness, mechanisms, and competitive influence. *NBER Working Paper No. 28477*.

Dahabreh, I. J. and M. A. Hernán (2019). Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology 34*, 719–722.

DeGolia, A. H., E. H. Hiroyasu, and S. E. Anderson (2019). Economic losses or environmental gains? Framing effects on public support for environmental management. *PLoS One 14*(7), e0220320.

Degtiar, I. and S. Rose (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application 10*, 501–524.

Della Valle, N. and P. Bertoldi (2021). Mobilizing citizens to invest in energy efficiency.

Domurat, R., I. Menashe, and W. Yin (2021). The role of behavioral frictions in health insurance marketplace enrollment and risk: Evidence from a field experiment. *American Economic Review 111*(5), 1549–1574.

Dörre, K. (2015). Tests for the underclass: The social effects of activating labour market policy in Germany. In *The New Social Division: Making and Unmaking Precariousness*, pp. 83–100. Springer.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics 4*, 3895–3962.

Duma, N., J. Vera Aguilera, J. Paludo, C. L. Haddox, M. Gonzalez Velez, Y. Wang, K. Leventakos, J. M. Hubbard, A. S. Mansfield, R. S. Go, et al. (2018). Representation of minorities and women in oncology clinical trials: Review of the past 14 years. *Journal of Oncology Practice 14*(1), e1–e10.

Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association 15*(3), 692–719.

Escobar, L. and A. Pedraza (2023). Active trading and (poor) performance: The social transmission channel. *Journal of Financial Economics 150*(1), 139–165.

Essien, U. R., S. B. Dusetzina, and W. F. Gellad (2021). A policy prescription for reducing health disparities—achieving pharmacoequity. *Jama 326*(18), 1793–1794.

Fang, X., L. Goette, B. Rockenbach, M. Sutter, V. Tiefenbeck, S. Schoeb, and T. Staake (2023). Complementarities in behavioral interventions: Evidence from a field experiment on resource conservation. *Journal of Public Economics 228*, 105028.

Federal Statistical Office of Germany (2023a). Haushalte nach Haushaltsgröße und Haushaltsmitgliedern. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Haushalte-Familien/Tabellen/1-2-privathaushalte-bundeslaender.html. Last retrieved: 11/21/2023.

Federal Statistical Office of Germany (2023b). Tabelle 61243-0001: Strompreise für Haushalte: Deutschland, Halbjahre, Jahresverbrauchsklassen, Preisarten.

Federal Statistical Office of Germany (2023c). Umweltökonmische Gesamtrechnungen. Stromverbrauch der privaten Haushalte nach Haushaltsgrößenklassen. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/private-haushalte/Tabellen/stromverbrauch-haushalte.html. Last retrieved: 10/30/2023.

Federal Statistical Office of Germany (2023d). Wohnfläche. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Wohnen/Tabellen/tabelle-wo4-wohnflaeche.html. Last retrieved: 11/21/2023.

Fehr, D., G. Fink, and B. K. Jack (2022). Poor and rational: Decision-making under scarcity. *Journal of Political Economy 130*(11), 2862–2897.

Finkelstein, A. and M. J. Notowidigdo (2019). Take-up and targeting: Experimental evidence from SNAP. *The Quarterly Journal of Economics 134*(3), 1505–1556.

Fowlie, M., M. Greenstone, and C. Wolfram (2015). Are the non-monetary costs of energy efficiency investments large? Understanding low take-up of a free energy efficiency program. *American Economic Review 105*(5), 201–204.

Fowlie, M., M. Greenstone, and C. Wolfram (2018). Do energy efficiency investments deliver? Evidence from the Weatherization Assistance Program. *The Quarterly Journal of Economics 133(3)*, 1597–1644.

Francesconi, M. and J. James (2021). None for the road? Stricter drink driving laws and road accidents. *Journal of Health Economics 79*, 102487.

Gal, D. and D. D. Rucker (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology 28*(3), 497–516.

Goldstein, B. A., M. Phelan, N. J. Pagidipati, R. R. Holman, M. J. Pencina, and E. A. Stuart (2019). An outcome model approach to transporting a randomized controlled trial results to a target population. *Journal of the American Medical Informatics Association 26*(5), 429–437.

Gonzales, M. H., E. Aronson, and M. A. Costanzo (1988). Using social cognition and persuasion to promote energy conservation: A quasi-experiment. *Journal of Applied Social Psychology 18*(12), 1049–1066.

Gopalan, M. and M. A. Pirog (2017). Applying behavioral insights in policy analysis: Recent trends in the United States. *Policy Studies Journal 45*(S1), S82–S114.

Gravert, C. A. (2021). Reminders as a tool for behavior change. *Published in Behavioral Science in the Wild, Toronto, Canada: University of Toronto Press*.

Guyton, J., D. S. Manoli, B. Schafer, and M. Sebastiani (2016). Reminders & recidivism: Evidence from tax filing & EITC participation among low-income nonfilers.

Hallsworth, M. (2023). A manifesto for applying behavioural science. *Nature Human Behaviour 7*(3), 310–322.

Halpern, D. and M. Sanders (2016). Nudging by government: Progress, impact & lessons learned. *Behavioral Science & Policy 2*(2), 53–65.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009—1055.

Harrison, G. W., K. Morsink, and M. Schneider (2020). Do no harm? The welfare consequences of behavioural interventions. *CEAR Working Paper 2020*.

Haushofer, J. and E. Fehr (2014). On the psychology of poverty. *Science 344*(6186), 862–867.

Holzmeister, F., J. Huber, M. Kirchler, and R. Schwaiger (2022). Nudging debtors to pay their debt: Two randomized controlled trials. *Journal of Economic Behavior & Organization 198*, 535–551.

Homar, A. R. and L. K. Cvelbar (2021). The effects of framing on environmental decisions: A systematic literature review. *Ecological Economics 183*, 106950.

Hotard, M., D. Lawrence, D. D. Laitin, and J. Hainmueller (2019). A low-cost information nudge increases citizenship application rates among low-income immigrants. *Nature Human Behaviour 3*(7), 678–683.

Hotz, V. J., G. W. Imbens, and J. H. Mortimer (2005). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics 125*(1-2), 241–270.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). Getting to the top of mind: How reminders increase saving. *Management Science 62*(12), 3393–3411.

Khanna, T. M., G. Baiocchi, M. Callaghan, F. Creutzig, H. Guias, N. R. Haddaway, L. Hirth, A. Javaid, N. Koch, S. Laukemper, et al. (2021). A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings. *Nature Energy 6*(9), 925–932.

Köster, T. (2023). *Arbeitslosigkeit unter inklusionstheoretischer Betrachtung nach Phelps: Eine Fallstudie*, Volume 19. Walter de Gruyter GmbH & Co KG.

Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes 75*(1), 23–55.

Laibson, D. and J. A. List (2015). Principles of (behavioral) economics. *American Economic Review 105*(5), 385–390.

Linos, E., A. Prohofsky, A. Ramesh, J. Rothstein, and M. Unrath (2022). Can nudges increase take-up of the EITC? Evidence from multiple field experiments. *American Economic Journal: Economic Policy 14*(4), 432–452.

List, J. A. (2020). Non est disputandum de generalizability? A glimpse into the external validity trial. *NBER Working Paper w27535*.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

Löschel, A., M. Rodemeier, and M. Werthschulte (2023). Can self-set goals encourage resource conservation? Field experimental evidence from a smartphone app. *European Economic Review 160*, 104612.

Marchionni, C. and S. Reijula (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science Part A 73*, 54–63.

Mullainathan, S. and E. Shafir (2013). *Scarcity: Why having too little means so much*. Macmillan.

Park, J., W. Son, H. Moon, and J. Woo (2023). Nudging energy efficiency behavior: The effect of message framing on implicit discount rate. *Energy Economics 117*, 106485.

Pearl, J. and E. Bareinboim (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 25, pp. 247–254.

Schulte, I. and P. Heindl (2017). Price and income elasticities of residential energy demand in Germany. *Energy Policy 102*, 512–528.

Shah, A. K., S. Mullainathan, and E. Shafir (2012). Some consequences of having too little. *Science 338*(6107), 682–685.

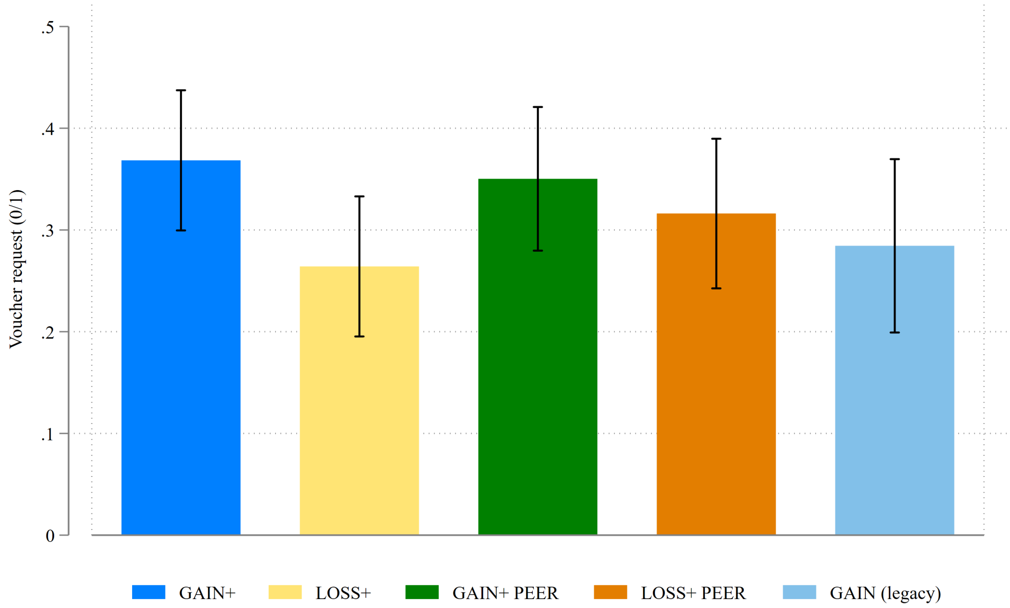
Stojanovski, O., G. W. Leslie, F. A. Wolak, J. E. H. Wong, and M. C. Thurber (2020). Increasing the energy cognizance of electricity consumers in Mexico: Results from a field experiment. *Journal of Environmental Economics and Management 102*, 102323.

Sussman, R., M. Chikumbo, and R. Gifford (2018). Message framing for home energy efficiency upgrades. *Energy and Buildings 174*, 428–438.

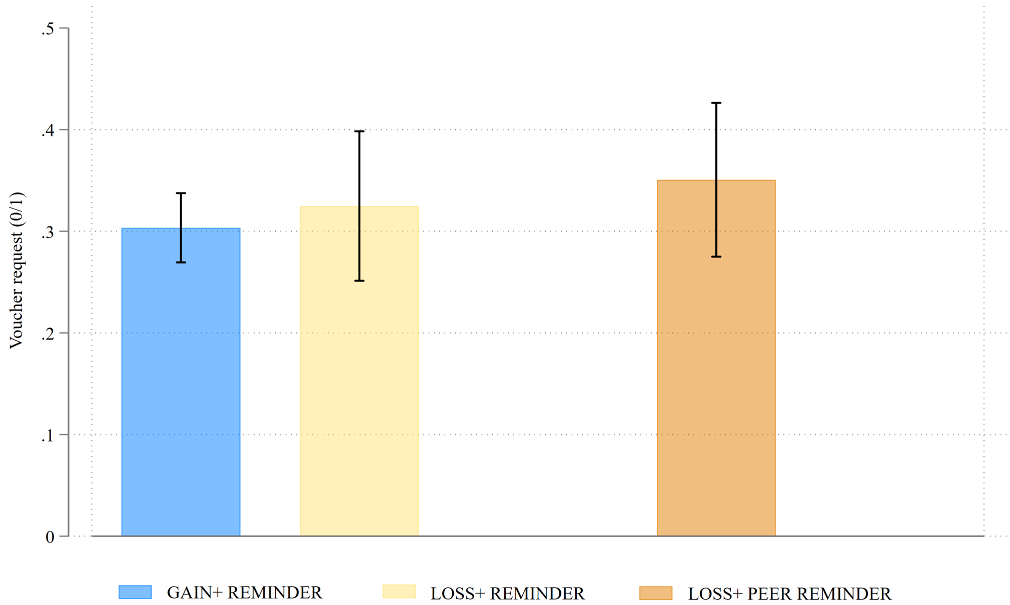
US Department of Housing and Urban Development (2024). The Green and Resilient Retrofit Program (GRRP). https://www.hud.gov/GRRP. Last retrieved: 08/01/2024.

Westreich, D., J. K. Edwards, C. R. Lesko, E. Stuart, and S. R. Cole (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology 186*(8), 1010–1014.

Appendix



(a) Information letter treatments.



(b) Reminder treatments.

Figure A.1: Voucher request rates by treatment group.

Table A.1: Treatment effects on refrigerator replacement conditional on having requested the voucher

	(1)	(2)	(3)	(4)
	Refrigerator replacement (0/1)			
$GAIN^{+}$	REF	REF	REF	REF
$LOSS^{+}$	-0.132	-0.167*	-0.144	-0.197*
	(0.095)	(0.094)	(0.108)	(0.111)
$GAIN^{+}$ PEER	-0.179*	-0.245**	-0.167	-0.229*
	(0.091)	(0.100)	(0.103)	(0.120)
$LOSS^{+}$ PEER	-0.257***	-0.247**	-0.256**	-0.228*
	(0.097)	(0.104)	(0.121)	(0.126)
GAIN (legacy)	0.043	0.009	-0.052	-0.071
	(0.100)	(0.098)	(0.117)	(0.115)
$GAIN^{+}$ REMINDER	-0.084	-0.108	-0.132*	-0.135*
	(0.066)	(0.069)	(0.073)	(0.076)
$LOSS^{+}$ REMINDER	-0.122	-0.136	-0.130	-0.128
	(0.090)	(0.092)	(0.092)	(0.097)
$LOSS^{+}$ PEER REMINDER	-0.154	-0.196*	-0.210*	-0.249**
	(0.094)	(0.100)	(0.110)	(0.117)
Constant	0.754***	0.423**	0.741***	0.460*
	(0.072)	(0.168)	(0.086)	(0.249)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
Request=1	Yes	Yes	Yes	Yes
N	571	555	541	525

Note: Linear probability models of refrigerator replacement (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The $GAIN^{+}$ treatment is the omitted reference treatment group. All regressions control for the communicated savings from replacement. Columns (2) and (4) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state, the social benefit transfer scheme, and whether the household heats warm water with electricity. Columns (3) and (4) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

Table A.2: Treatment effects on voucher request: Allowing for differential reminder effects

	(1)	(2)	(3)	(4)
	Voucher request (0/1)			
$GAIN^{+}$	REF	REF	REF	REF
$LOSS^{+}$	-0.105**	-0.114**	-0.108**	-0.112**
	(0.050)	(0.049)	(0.047)	(0.047)
$GAIN^{+}$ PEER	0.031	0.056	0.000	-0.004
	(0.053)	(0.054)	(0.055)	(0.057)
$LOSS^{+}$ PEER	-0.004	0.032	0.003	0.014
	(0.054)	(0.055)	(0.053)	(0.053)
GAIN (legacy)	-0.098*	-0.085	-0.078	-0.072
	(0.056)	(0.055)	(0.054)	(0.053)
$GAIN^{+}$ Letter/SMS	-0.049	-0.042	-0.050	-0.036
	(0.043)	(0.043)	(0.041)	(0.041)
$GAIN^{+}$ Tag	-0.061	-0.059	-0.068	-0.052
	(0.049)	(0.049)	(0.046)	(0.046)
$GAIN^{+}$ Letter/SMS Tag	-0.099**	-0.100**	-0.103**	-0.109**
	(0.048)	(0.048)	(0.046)	(0.046)
$LOSS^{+}$ Tag	-0.044	-0.048	-0.044	-0.042
	(0.051)	(0.050)	(0.047)	(0.047)
$LOSS^{+}$ PEER Tag	0.031	0.071	-0.033	-0.034
	(0.055)	(0.056)	(0.055)	(0.056)
Constant	0.301***	0.540***	0.310***	0.419***
	(0.043)	(0.085)	(0.045)	(0.101)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
N	1802	1761	1785	1745

Note: Linear probability models of voucher request (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The $GAIN^{+}$ treatment is the omitted reference treatment group. All regressions control for the communicated savings from replacement. Columns (2) and (4) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state, the social benefit transfer scheme, and whether the household heats warm water with electricity. Columns (3) and (4) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

Table A.3: Treatment effects on refrigerator replacement: Allowing for differential reminder effects

	(1)	(2)	(3)	(4)
	Refrigerator replacement (0/1)			
GAIN^{+}	REF	REF	REF	REF
LOSS^{+}	-0.104**	-0.117***	-0.102**	-0.113***
	(0.042)	(0.041)	(0.040)	(0.040)
GAIN^{+} PEER	-0.044	-0.045	-0.062	-0.075
	(0.046)	(0.046)	(0.046)	(0.046)
LOSS^{+} PEER	-0.088*	-0.064	-0.071	-0.056
	(0.045)	(0.046)	(0.046)	(0.046)
GAIN (legacy)	-0.051	-0.051	-0.046	-0.047
	(0.049)	(0.048)	(0.048)	(0.047)
GAIN^{+} Letter/SMS	-0.081**	-0.081**	-0.082**	-0.072**
	(0.037)	(0.037)	(0.036)	(0.036)
GAIN^{+} Tag	-0.030	-0.037	-0.041	-0.031
	(0.043)	(0.043)	(0.042)	(0.042)
GAIN^{+} Letter/SMS Tag	-0.088**	-0.093**	-0.095**	-0.104***
	(0.041)	(0.041)	(0.040)	(0.040)
LOSS^{+} Tag	-0.070	-0.078*	-0.081*	-0.084**
	(0.043)	(0.044)	(0.042)	(0.042)
LOSS^{+} PEER Tag	-0.035	-0.020	-0.090*	-0.097**
	(0.047)	(0.048)	(0.046)	(0.047)
Constant	0.234***	0.292***	0.233***	0.239**
	(0.037)	(0.072)	(0.038)	(0.095)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
N	1802	1761	1785	1745

Note: Linear probability models of refrigerator replacement (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The GAIN^{+} treatment is the omitted reference treatment group. All regressions control for the communicated savings from replacement. Columns (2) and (4) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state, the social benefit transfer scheme, and whether the household heats warm water with electricity. Columns (3) and (4) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

Table A.4: Heterogeneous treatments effects on refrigerator replacement by transfer type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Refrigerator replacement (0/1)							
GAIN$^+$	REF	REF	REF	REF	REF	REF	REF	REF
LOSS$^+$	-0.010	-0.028	-0.030	-0.035	-0.159***	-0.168***	-0.130***	-0.141***
	(0.079)	(0.078)	(0.081)	(0.081)	(0.047)	(0.047)	(0.046)	(0.047)
GAIN$^+$ PEER	-0.070	-0.024	-0.128	-0.088	-0.013	-0.057	0.024	-0.024
	(0.080)	(0.081)	(0.087)	(0.090)	(0.056)	(0.056)	(0.057)	(0.059)
LOSS$^+$ PEER	-0.010	0.042	0.019	0.062	-0.104**	-0.115**	-0.089*	-0.100*
	(0.088)	(0.088)	(0.093)	(0.097)	(0.052)	(0.053)	(0.054)	(0.054)
GAIN (legacy)	0.071	0.088	0.072	0.083	-0.141***	-0.139**	-0.108**	-0.112**
	(0.089)	(0.087)	(0.103)	(0.104)	(0.055)	(0.055)	(0.055)	(0.055)
GAIN$^+$ REMINDER	-0.014	-0.009	-0.033	-0.021	-0.100**	-0.104**	-0.083**	-0.083**
	(0.060)	(0.059)	(0.062)	(0.065)	(0.042)	(0.043)	(0.040)	(0.040)
LOSS$^+$ REMINDER	-0.063	-0.054	-0.108	-0.100	-0.076	-0.081	-0.064	-0.068
	(0.074)	(0.074)	(0.077)	(0.082)	(0.054)	(0.055)	(0.053)	(0.053)
LOSS$^+$ PEER REMINDER	0.025	0.051	0.012	0.017	-0.047	-0.056	-0.086	-0.104*
	(0.087)	(0.087)	(0.096)	(0.101)	(0.056)	(0.057)	(0.054)	(0.055)
Constant	0.236***	0.474***	0.245***	0.332**	0.207***	0.227***	0.185***	0.249*
	(0.071)	(0.132)	(0.079)	(0.152)	(0.045)	(0.086)	(0.047)	(0.127)
ALG II	No	No	No	No	Yes	Yes	Yes	Yes
Savings Info	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
N	638	620	610	591	1164	1141	1129	1105

Note: Linear probability models of refrigerator replacement (yes/no) on treatments. The treatments are included as indicators for the respective treatment group. The GAIN$^+$ treatment is the omitted reference treatment group. Columns (1)-(4) cover program participants who receive social transfer payments other than long-term unemployment benefits (ALG II=no). Columns (5)-(8) cover program participants receiving long-term unemployment benefits (ALGI II=yes). All regressions control for the communicated savings from replacement. Columns (2), (4), (6) and (8) add control variables for household's electricity price, number of persons in the household, past electricity consumption, living space, federal state and whether the household heats warm water with electricity. Columns (3), (4), (7) and (8) include fixed effects for the respective intervention sites, the month of when the participant is informed about his/her replacement eligibility and a month-site interaction. Robust standard errors are in parenthesis. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.