UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Transporting behavioral insights to low-income households: a field experiment on energy efficiency investments

Bettina Chlond, Timo Goeschl, Martin Kesternich, Madeline Werthschulte

Angaben zur Veröffentlichung / Publication details:

Chlond, Bettina, Timo Goeschl, Martin Kesternich, and Madeline Werthschulte. 2026. "Transporting behavioral insights to low-income households: a field experiment on energy efficiency investments." *Journal of Public Economics* 263: 105780. https://doi.org/10.1016/j.jpubeco.2026.105780.

Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY 4.0: Creative Commons: Namensnennung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by/4.0/deed.de

Journal of Public Economics 263 (2026) 105780



Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/pubec



Transporting behavioral insights to low-income households: A field experiment on energy efficiency investments☆

Bettina Chlond[a,e], Timo Goeschl[b,e], Martin Kesternich[c,e,*], Madeline Werthschulte[d,e]

[a] ETH Zurich, Zurich, Switzerland
[b] Heidelberg University, Heidelberg, Germany
[c] University of Augsburg, Augsburg, Germany
[d] Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[e] ZEW – Leibniz Centre for European Economic Research, Mannheim, Germany

HIGHLIGHTS

- Behavioral insights may not generalize to under-represented target populations.
- We implement an RCT in one of the largest energy efficiency assistance programs.
- Most behavioral insights proved ineffective or inefficient.
- Loss-framing backfired by reducing take-up.
- We rank insights by their Marginal Value of Public Funds (MVPF).

ARTICLE INFO

JEL classification:
C93
D91
Q49

Keywords:
External validity
Low-income households
Field experiment
Randomized controlled trial
Governmental welfare programs
Energy efficiency
Technology adoption

ABSTRACT

Established behavioral insights may not generalize when applied to under-represented target populations. Iterative approaches that identify such external validity failures to guide policy deployment at scale and to inform subsequent design refinements gain importance. We implement this approach in a field experiment involving 1800 low-income households in one of the largest energy efficiency assistance programs. Despite extensive co-designed vetting, most behavioral insights proved ineffective or inefficient. Notably, loss-framed interventions backfired by reducing take-up. We rank insights by their Marginal Value of Public Funds (MVPF), identifying 'unexpected winners' that outperform alternatives rated as promising by experts and laypeople. Our analysis suggests that factors specific to low-income households, such as scarcity, fundamentally alter behavioral responses, necessitating guided deployment to optimize future public assistance design.

1. Introduction

"Behavioral insights" such as loss framing, reminders or peer comparisons have increasingly been recommended to (OECD, 2025; Hallsworth, 2023) and deployed by policy-makers in industrialized countries because they promise to combine reliable effectiveness, low demands on public budgets, and high acceptance among the public (Halpern and Sanders, 2016; Gopalan and Pirog, 2017). At the same time, a growing body of work cautions that such insights may not be externally valid across environments and populations, in particular when the target group was systematically under-represented or even absent from the evidence base on which the insights were built (Bryan et al., 2021; Hallsworth, 2023). For program managers considering whether to deploy a well-established insight in their context, this raises at least three practical tasks: detecting problems of external validity before they affect the target population at scale, guiding the program design away

☆ The experiment was pre-registered at the AEA RCT Registry as trial AEARCTR-0010757. The research has been funded by the Fritz Thyssen Foundation (grant number 10.21.2.028WW), IRB approval reference number: FESS20224016.

* Corresponding author at: University of Augsburg, Augsburg, Germany.
Email address: martin.kesternich@uni-a.de (M. Kesternich).

https://doi.org/10.1016/j.jpubeco.2026.105780
Received 27 February 2025; Received in revised form 30 July 2026; Accepted 8 September 2026
Available online 24 September 2026

0047-2727/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Open access funding enabled and organized by Projekt DEAL.

from detected failures, and learning from such failures for future policy design.

This paper reports on a co-designed randomized controlled trial (RCT) of behavioral insights carried out over a year as part of an overhaul of one of the world's largest energy efficiency assistance schemes. The target group of this scheme – with an annual budget exceeding 12 million euros – are recipients of federal income support in Germany. The objectives and constraints of the policy overhaul are common in public policy. The objective was to improve the effectiveness and efficiency of a core part of the scheme, the so-called "Refrigerator Replacement Program" (RRP). The four main constraints were: budgetary, necessitating low- or zero-cost changes; practical, requiring changes scalable across 150 project sites; political, precluding changes that undermine a redistributive climate policy measure; and ethical, ruling out changes that are harmful to the vulnerable target group.

In this setting, we present an RCT-based approach for detecting problems of external validity of program alternatives within this program and for guiding at-scale deployment of those alternatives that deliver detectable improvements in both effectiveness and efficiency. In this "guided approach", co-designed behavioral insights are not only screened for compliance with the SANS conditions[1] for scalability (List, 2020) but also randomly administered to target subjects over a meaningful period. They are additionally evaluated ex-post for effectiveness and ranked by their change in social welfare, using the marginal value of public funds (MVPF) as an efficiency metric (Hahn et al., 2026). This sequence of steps – pre-screening, piloting, and ex-post evaluation prior to scale-up – contrasts with a more conventional, "unguided" deployment path, in which a behavioral insight with a strong track record elsewhere is adopted directly on the strength of that track record. In our context, this guided approach resulted in the random assignment – over the course of one year – of target households from a random sample of program sites to three behavioral insights (loss framing, reminders, peer experience) in eight treatment arms, testing for treatment effects and assessing their MVPFs.

The co-design process underpinning our RCT engaged policy-makers and built on several blocks of prior knowledge at the pre-screening stage. First, it relied on existing evidence of behavioral insights and their applicability to Germany, ranging from meta-studies of multi-country evidence (Khanna et al., 2021; Composto and Weber, 2022) to household investments in energy efficiency durables (Silvi and Padilla Rosa, 2021; Park et al., 2023) and voucher-based public policy programs (Banerjee et al., 2023; Bergman et al., 2024). Second, it drew on program evidence that small procedural changes to the RRP were capable of delivering measurable improvements in performance (Chlond et al., 2025). Third, it integrated tacit and explicit practitioner knowledge elicited through a series of workshops with local site managers to jointly identify suitable interventions, thus responding to concerns over poor external validity threatening the ethical and political constraints of the program. Support for such concerns came from prior evidence (e.g., Bronchetti et al., 2013; Linos et al., 2022; Domurat et al., 2021; Holzmeister et al., 2022) that demonstrated that low-income households respond differently to seemingly well-established behavioral insights, sometimes even in the opposite direction. Fourth, we built on predictions both by experts and by members of the general public who were asked to vet two of the program alternatives for expected impacts. The result of this process was three behavioral insights to be tested as potential solutions that also respected the budgetary and practical constraints.

Using this guided approach in our context, we find an unexpectedly low external validity of common behavioral insights: the majority of program alternatives fail to deliver improvements both in effectiveness and efficiency relative to the management baseline. In fact, despite extensive vetting through co-design, some behavioral insights not only fail to deliver improvements, but actually backfire. In our RCT, for example, loss framing (Kühberger, 1998; Homar and Cvelbar, 2021), widely regarded as relying on a 'behavioral principle' (OECD, 2025) and predicted to raise performance, leads to significantly worse effectiveness than the management baseline. Reminders, another mainstay of behavioral insights (Guyton et al., 2016; Gravert, 2021), similarly reduce effectiveness. Given their demonstrated effectiveness in similar circumstances, geographical and otherwise, their present performance does not conclusively prove failure, but rather highlights the risk of their at-scale deployment among underrepresented target groups. In such a setting, an unguided deployment which neither screens co-designed behavioral insights for their scalability nor evaluates them ex-post with respect to their effectiveness faces the risk of violating ethical constraints (Barrett et al., 2019). Beyond questions of effectiveness, an unguided deployment can also violate political or welfarist constraints that take into account the efficient use of public funds dedicated to the program. This is the case when policy alternatives fail to deliver compensating improvements in effectiveness in return for higher program costs. In our context, this became evident in the low MVPF of reminders as a policy alternative, which reflected not just low effectiveness, but also increased cost of sending reminders. Taken together, our evidence cautions against unguided deployment of established behavioral interventions in ethically problematic and politically charged contexts while our approach could provide a possible template for a guided policy overhaul.

Corollary evidence from the RCT supports learning for future policies by providing insights into the channels that caused loss framing and reminders to fail in our setting. First, on loss framing, our evidence 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 those used by the average household (Kőszegi and Rabin, 2006). It is consistent with evidence that loss tolerance is more prevalent among those with experience of adverse financial shocks (Chapman et al., 2025). It is also consistent with arguments that households experiencing scarcity display more risk-seeking and impatient attitudes (Haushofer and Fehr, 2014; Mullainathan and Shafir, 2013). Second, the evidence on reminders is consistent with the idea that the announcement of a reminder leads to lower task completion because anticipated reminders undermine one's own investment in imperfect memory (Ericson, 2017). This may be particularly relevant to cognitively stressed low-income households.

The paper is structured as follows. Section 2 situates our inquiry in the existing literature and details its contributions. Section 3 describes our experimental design, providing detailed information on the procedures of the energy efficiency assistance scheme, the treatment development and rationale, and the rollout of the RCT. We next describe our household sample in Section 4 and our empirical strategy in Section 5. Section 6 contains the key results on treatment effects that support the ranking of design alternatives based on their MVPF in Section 7. Section 8 concludes.

2. Literature and contribution

Our study lies at the intersection of four strands of literature. First, it contributes to an emerging literature on the effects of behavioral interventions on the performance of governmental welfare and public assistance programs. Much of this literature concerns questions of effectiveness, for example regarding the 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), medicare insurance (Brot-Goldberg et al., 2023) or energy subsidies (Hahn and Metcalfe, 2021). A common finding is low effectiveness: take-up rates are usually rather modest. For example, only

[1] These conditions impose a burden of proof on the experimenter with respect to selection, attrition, naturalness, and scaling of the experiment.

14 percent claim the EITC benefits (Bhargava and Manoli, 2015), 6 percent of eligible low-income households enroll in the SNAP (Finkelstein and Notowidigdo, 2019) and 4.3 percent claim energy subsidies (Hahn and Metcalfe, 2021).

At the same time, this literature reports mixed evidence on the effectiveness of different behavioral interventions on take-up among low-income households, including presentational changes in program descriptions in letters and reminders about the program. On the one hand, 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 when the information letter further includes assistance information. Looking at the impact of different mail reminders on tax compliance, Hallsworth et al. (2017) report that the reminders increase tax compliance among delayed taxpayers. While there is no significant difference between gain- and loss-framed reminders in the whole sample, loss-framed messages have a particularly strong impact for those with the largest debt. On the other hand, Hahn and Metcalfe (2021) investigate the impact of different information letters on enrollment rates in an energy subsidy program. While receiving a letter increases enrollment rates significantly, the particular behavioral content of the letters appears to be less important with, e.g., no differences in enrollment rates between a gain and a loss frame. Likewise, 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. Recently, this literature has extended beyond effectiveness to examine welfare through the lens of the marginal value of public funds spent on these programs (Hendren and Sprung-Keyser, 2020; Hahn et al., 2026). Using the MVPF metric forces the researcher to investigate the additional cost of the policy that has generated the observed effects, with potentially large impact on how policies are ranked.

We contribute to this literature by offering new international evidence on a public assistance program from an OECD country that incentivizes investment decisions rather than providing support in cash or in kind. While these investments have potential for sustained improvements in household finances, they also pose particular challenges for low-income households by requiring cash to be diverted from consumption to investment and due to the behavioral challenges specific to optimization over a longer time horizon.

Second, the paper contributes to the literature on the external validity of behavioral insights, in particular in public policy (Pearl and Bareinboim, 2011; Allcott, 2015; Degtiar and Rose, 2023). One manifestation occurs 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 external validity cannot be assumed. One threat arises when the study sample is not a subset of the target population (Duflo et al., 2007). 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 limited external validity (Barrett et al., 2019). We contribute to this literature by conducting an RCT in which we test behavioral insights that have consistently performed successfully in the general population. By showing that applying some of these insights to a new target population not only fails to improve performance, but can also lower it, we demonstrate the importance of heterogeneity in behavioral public policy.

The third strand of literature to which we contribute deals with the impact of behavioral interventions on energy savings and efficiency among 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 about the effectiveness of behavioral interventions on investments by low-income households. One particular reason is that empirical evidence from non-targeted energy savings programs provides 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 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 ($1000 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 of the external validity of behavioral interventions to new study populations. For example, 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.

Finally, our study relates to a recent interest within public economics in how to combine existing evidence, prior beliefs and purposeful experimentation to best inform policy design before scaling (List, 2020, 2024; Athey, 2025). We contribute to this idea of adaptive public policy design through an efficacy trial that implements this practice for a nationally consequential policy. This study exemplifies how governments could institutionalize iterative, evidence-based optimization in redistributive climate policy without compromising scalability or equity objectives.

3. Experimental design

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

The "Energy-Saving-Check" (SSC, German: *Stromspar-Check*) is the largest nation-wide energy efficiency assistance scheme in Germany, that aims to lower the energy bills of low-income households 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 12 million euros 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, since 2009 the Refrigerator Replacement Program (RRP; German: *Kühlgeräte-Tauschprogramm*) has been offering cash vouchers to households on federal income support[2] in order to encourage households to replace their old and inefficient refrigeration devices with modern, highly efficient models. The rationale for enrolling households in the

[2] 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

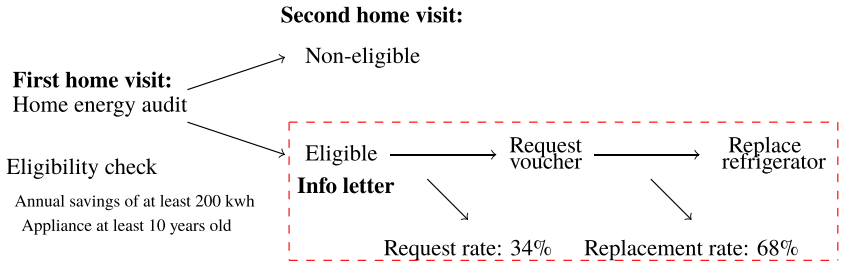


Fig. 1. Procedure of the home audits.

RRP is the large contribution, roughly 25 percent (Bundesverband der Energie- und Wasserwirtschaft e.V. (BDEW), 2019), that refrigerators make to the electricity consumption of the average German household.[3] Differences in refrigerator efficiency can therefore significantly impact residential electricity bills.

For recruiting qualified households into the SSC's home energy audit, the program is actively promoted in employment and social assistance agencies through both printed and audiovisual materials. Some local branches of the social assistance agency mandate the participation of households with excessively high energy bills. In addition, active staffers provide individualized descriptions of the program using pop-up booths in shopping streets and malls. 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 (see Fig. 1 for a summary of the procedure). 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. To be eligible for enrollment, 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 classes of devices on the market.[4] The screening leads to differences in the second visit: Both eligible and non-eligible households receive an energy-saving kit.[5] 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 enrollment in the RRP through educational material and promotion.[6]

An important part of this material is an information letter (see Fig. A.1). It contains all relevant information for refrigerator replacement 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 euros) from successful replacement (see green box in Fig. A.1). Second, it offers a step-by-step explanation of 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 fulfills the requirements, (pre-)financing the purchase, and organizing the logistics of delivering the new and disposing of the old refrigerator. Once requested, the voucher is valid for two months. Consequently, there exists a sharp 'deadline' for 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 Fig. 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 refrigerators. The overall replacement rate of 23 percent consequently falls substantially short of the 100 percent replacement target that the program designers aim for given the cost-benefit evaluation of the investment that every eligible household has passed.[7]

3.2. Treatments

Our study design builds on several blocks of prior knowledge. One source is the insights from behavioral economics. The second source is the results of our own empirical research on the impacts of past program changes on the performance of the RRP. Prior own evidence from the RRP showed that the impacts of program changes on program performance are sometimes hard to predict (Chlond et al., 2025): On the one hand, program take-up among low-income households was higher for higher levels of financial assistance. This conformed to intuitive expectations of both researchers and program staff. On the other hand, program take-up was higher when assistance procedures relied on elective (opt-in) rather than default enrollment and imposed rigid deadlines for replacement. This finding was unexpected and highlighted that underappreciated behavioral mechanisms can lead to surprising performance effects. The third source is the views of program officials and local site managers who were engaged in the co-design process. The co-design process with program officials and local site managers comprised the joint development and pre-screening of a set of treatment variations within or alongside the information letter. The aim of the process was to ensure effortless administration, high naturalness, and full scalability to all program sites in the country.[8] The fourth source of

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).

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

[4] 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 belongs to energy class D or higher on the EU Energy Label scale. The investment of *every* eligible household passes a cost-benefit evaluation.

[5] 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 advisors.

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

[7] The average new appliance purchased in the program costs 557 euros. Including the federal subsidy of 100 euros that all participants receive, the investment pays back after 4.1 to 4.7 years at respective discount rates of 0 percent and 5 percent. In addition, further state and local subsidy programs top up the federal subsidy at several branches. This further shortens the pay-back time.



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

(a) GAIN$^+$ Frame.



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!

(b) LOSS$^+$ Frame.

Fig. 2. Treatment variations: gain and loss frames.

priors is third-party beliefs about the likely performance improvements generated by two specific interventions through an incentivized prediction task in 2022. Two types of third parties were targeted, experts and members of the general public. The experts were recruited through a systematic collection of municipal social assistance officers' email addresses and through an email campaign among branch managers of the SSC program who were not working at treatment sites. This yielded 75 unique responses. The members of the general public were recruited via the research platform Prolific, with participation restricted to adults residing in Germany. This yielded 500 unique responses. Participants were invited to bet on the take-up rate achieved by the intervention compared to the rate achieved by the baseline, with rewards of up to 5 euros to be paid upon completion of the intervention.

Our interventions target three main dimensions. First, reported annual savings in electricity costs can be framed as a financial GAIN or as a financial LOSS if households forgo the replacement opportunity. The framing dimension is captured in our first three treatments. The GAIN treatment simply reflects a legacy version of the information letter (see Fig. A.1). The 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 Fig. 2a) 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 Fig. 2b). 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) or from actual replacements recently conducted in households with similar characteristics, specifically the household composition.[9] The engineering estimate of savings reports 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. To realize these financial gains, it is assumed that households exactly follow 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. The 'peer experience' of savings, by contrast, reports the realized monetary values based on actual replacements by a peer group with a similar household composition.[10]

We highlight the reference to peer-behavior in two ways. First, we add a pictogram of the respective household composition on the right side of the savings information. 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 Fig. 3a). The alterations are similar in the loss frames (see Fig. 3b).

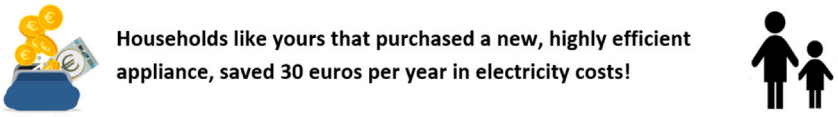
As a third treatment dimension, we introduce different reminders for eligible households (see Fig. 4). In a first variant, households were informed that they would be recontacted 4–8 weeks after receiving the information letter and asked about their preference for being contacted via SMS or via letter. To our knowledge, every household provided consent to be contacted via SMS or letter such that there is no self-selection into receiving a reminder. Conditional on households' preferences, local site managers then sent out a letter- or SMS-based reminder at the beginning of a new month (when households usually are more financially liquid; see Fig. 4a for a translation of the reminder text). In a second variant, the energy advisor placed a tag, which displays the logo of the Energy-Saving-Check program, inside the refrigerator during the second visit (see Fig. 4b).[11] In contrast to the information letter variations, the

[8] We started with a set of interactive online workshops where we jointly identified potential barriers to accessing the most economically relevant information provided in the letter and discussed a first set of potential interventions. One result of this process was an emphasis on sensitive language; 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. 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.

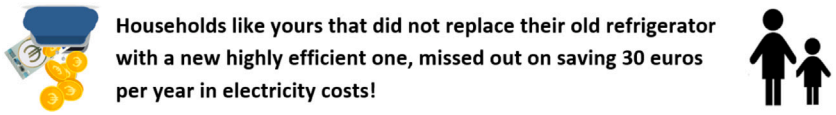
[9] Previous research has shown that even for the same model, engineering estimates and realized savings can differ substantially, with engineering estimates typically overstating realized savings (Davis et al., 2014; Goeschl, 2019).

[10] We calculate the annual savings realized by peers 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 euros savings); two adults (29 euros savings); two adults and one child (30 euros savings); two adults and two children (36 euros savings); two adults and three or more children (41 euros savings); single parenting (30 euros savings).

[11] 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 at low-income households.

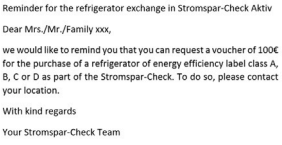


(a) GAIN+ PEER Frame.



(b) LOSS+ PEER Frame.

Fig. 3. Treatment variations: peer experience gain and loss frames.


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.

Fig. 4. Reminder treatments.

reminders required additional program resources to be spent, estimated at 10 euros per eligible household.

We organize these three treatment dimensions in a "2 × 3 + Reminder" design, as displayed in Fig. 5. 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 display the second treatment dimension, i.e., whether the savings are 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 gray fields in Fig. 5 display treatment versions that we test both with and without reminders.[12]

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 at the end of July 2022. To meet our scalability targets, the randomization process was carried out automatically by the program database. Households are allocated to the different treatments with equal probabilities, except for (6) GAIN+ REMINDER to which we over-sampled.[13]

3.3. Hypotheses

With our experimental design, we test four main hypotheses. Our first hypothesis concerns the impact of using visual enhancements or visual aids in an information letter about energy efficiency in order to reinforce the verbal message (Stojanovski et al., 2020; Allcott and Greenstone, 2024). 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.[14]

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

[12] Note, that this definition is agnostic with respect to the specific reminder types. Such differentiation between reminder types is, however, only relevant for the GAIN+ REMINDER treatment, which combines (i) the letter/SMS reminder, (ii) the refrigerator tag and (iii) both reminder types. In the loss domain, by contrast, both the loss frame and the peer-experience loss frame are combined with the refrigerator tag and kept separate (LOSS+ REMINDER and LOSS+ PEER REMINDER treatment). As a robustness check, we differentiate between the different reminder versions (see Tables B.7 and B.8). This procedure does not alter our main observations. We provide a detailed description of how our analysis in the paper has evolved compared to our PAP in Appendix D.

[13] The reason is that the GAIN+ Letter/SMS 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.

[14] A starting point is the seminal study by Gonzales et al. (1988) who examined the effect of exposing 408 homeowners who qualified for enrollment in an energy efficiency retrofit program to one of two different home energy audit procedures. In one, homeowners 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 enrollment among homeowners 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 of energy savings (see Homar and Cvelbar (2021) for a meta-study covering 61 studies). 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.

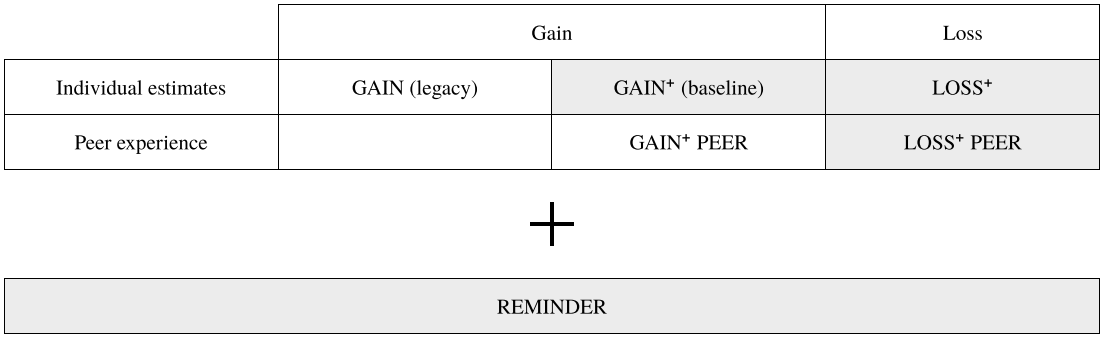


Fig. 5. Treatment dimensions.

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.

In our third-party belief elicitation, experts and the general public were asked to predict whether the LOSS$^+$ frame would lead to higher replacement rates than the legacy GAIN frame. Both predicted a significantly better performance, with experts predicting an average improvement of 3.2 percentage points in the replacement rate and the general public predicting an average improvement of 8.0 percentage points. These predictions broadly agree with the literature, but are weaker than H2 because the comparison group is the legacy design (GAIN) rather than the baseline (GAIN$^+$).

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 of framing is derived from the average household, with little discussion on whether the same patterns also apply to our specific low-income demographic.[15]

Thirdly, we hypothesize a reference to peer experiences to both reduce uncertainty about 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.[16] Hence, the peer experience may decrease the degree of uncertainty and increase the attachment that households place on the possible savings.

H3: Holding communicated savings fixed, refrigerator replacement rates will be higher in the peer reference design compared to a design that relies on individual projected estimates.

The fourth hypothesis focuses on the effect of reminders as a tool to overcome potential procrastination 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 among eligible households. Specifically, Chlond et al. (2025) 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 sizable effects. This also holds for the context of energy efficiency (Fang et al., 2023) and for 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 savings from replacing the appliance), and overemphasize the current cost (here, the outlay for replacing the appliance), leading them to postpone – and ultimately forgo – 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 than in the non-reminder treatments.

In our third-party belief elicitation, experts and the general public were asked to predict whether the tag reminder would lead to higher replacement rates compared to no reminder. Both predicted a significantly better performance, with experts predicting an average improvement of 8.3 percentage points in the replacement rate and the general public predicting an average improvement of 9.9 percentage points. These predictions again broadly agree with those in the literature.

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.[17] Two, the frame could determine the direction of the reminder effect, reinforcing or possibly reversing the effect.[18]

The other area of threat stems from asking households whether they prefer to be contacted via SMS or letter. Being aware that there will

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

[16] Likewise, research in other domains also 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.

[17] 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 at the time of reminder mailing and, hence, receives the reminder will not be the same across the two framings.

[18] For example, if H2 does not hold and households are discouraged by a loss frame, then a household can make sure that they 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 incentive for replacement.

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

3.4. Selection and training of sites

To address potential concerns about site selection and to support the scaling idea of our interventions, we randomly selected a number of local sites that were invited to participate in our field experiment. To this end, we collected data on the number of long-term unemployment benefit recipients in the 158 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 from the program officials to participate in our study in December 2021. 21 out of the 30 sites accepted our invitation.[19] After the selection process was 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 managers' choices. 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.[20] 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. The roll-out started in April 2022. We summarize the timeline of the co-design, training and selection process in Table B.1.

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,[21] (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 1803 households that fulfill these criteria.

Table 1 displays summary statistics of household information recorded in the program database. The average electricity price in the billing period 2021/2022 was 33 cents/kWh aligning with the average nation-wide consumer electricity rates (Federal Statistical Office of Germany, 2023b). Electricity consumption is notably lower compared to national statistics. While our study participants consume on average 2926 kWh per year, the German annual average is 3383 kWh (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 of our sample. Ex-ante engineering estimates of annual energy savings from refrigerator replacement are on average 336.33 kWh. Together with the average electricity price, this maps to average annual financial savings of 112.74 euros. 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 (71.1 m^2 on average) is lower than the national average (96.2 m^2, Federal Statistical Office of Germany (2023d)). These differences relative 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 receives a basic pension (Grundsicherung) and 12 percent receive housing benefits (Wohngeld).

To check for successful randomization, we test the balance of covariates across treatments (see Table B.2). We test for each covariate whether the mean in the respective treatment arm is statistically significantly different from the mean across all other groups. Some covariates show small deviations from the overall mean in specific treatment arms but after applying the correction by List et al. (2019) for multiple hypothesis testing, only the difference in estimated energy savings in treatment groups GAIN and GAIN$^+$ REMINDER is marginally significant. In our preferred estimation specification, we control for all covariates.[22] In addition, we use location-by-month fixed effects to exploit the full estimation sample and to avoid small sample issues in single treatment cells of interest. As an additional robustness check, we report a specification with advisor fixed effects.

The number of observations per treatment is about equally split across groups except for two exceptions. First, the number of observations for the GAIN group is lower due to technical issues that caused a delay in implementation. Data collection for this group only started at the end of July 2022. Second, the number of observations is higher for the GAIN$^+$ REMINDER group, since we summarize three initially implemented treatment groups that vary by the reminder version and we oversampled the GAIN$^+$ SMS/Letter REMINDER group as explained in Section 3.2.

Fig. A.4 displays the mean and 95 percent confidence intervals of our main outcome, refrigerator replacement, by treatment group.

[19] The sites not signing up were Chemnitz, Delmenhorst, Essen, Recklinghausen, Moers and two sites in Dortmund. 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 also excluded from the study. Three additional sites (Jena, Weimar, Erfurt) as well as a pilot site (Frankfurt) showed interest and self-selected into the experiment. To avoid resulting selection effects, we exclude these sites from our analysis. 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. Fig. A.2 displays all program sites (pink), the sites participating in our study (red) and the excluded self-selected sites (violet).

[20] 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.

[21] 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.

[22] Without applying the multiple hypothesis correction, the price paid for electricity is significantly different from the overall mean for treatment arms GAIN, GAIN$^+$ and GAIN$^+$ PEER. Energy consumption turns out to be significant in GAIN$^+$ and estimated savings in GAIN and GAIN$^+$ REMINDER.

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	0.20	0.30	0.30	0.32	0.34	0.40	0.71	1803
Energy consumption (kWh)	2926.10	1708.17	0	1250.00	1712.00	2572.33	3700.00	4957.83	17501.44	1762
Electric water heating: (1 = yes)	0.37	0.48	0	0.00	0.00	0.00	1.00	1.00	1	1803
Annual est. savings (kWh)	336.33	153.37	11.55	204.67	228.68	293.64	414.62	547.56	1466.04	1802
No. persons (count)	2.78	1.88	1	1.00	1.00	2.00	4.00	5.00	11	1803
Living space (m^2)	71.08	27.60	20	43.00	52.80	65.00	83.00	107.00	300	1803
Unemployment benefits (1 = yes)	0.65	0.48	0	0.00	0.00	1.00	1.00	1.00	1	1803

Note: Displayed are summary statistics for the variables electricity price (in euro/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.

Replacement rates, i.e., the share of households that actually replaced their refrigerator out of all households who were eligible for and informed about the subsidized replacement opportunity, range between 14 and 24 percent. We observe the highest replacement rates among households in the baseline treatment GAIN$^+$. The LOSS$^+$ treatment shows the lowest replacement rates. Between these two fall replacement rates of the treatments using peer experience to encourage take-up, mirroring the pattern of gain versus loss framing, as well as the legacy design GAIN. As Fig. A.4(b) 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 at 17 percent below the rate of 24 percent in the GAIN$^+$ treatment. The voucher request rates range between 26 and 37 percent.[23] As in the second stage, request rates are 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.

5. Empirical strategy

To analyze the effect of the 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:

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

where i denotes the individual household and Y_i refers to the 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 Eq. (1) as a linear probability model (LPM).

We define the GAIN$^+$ treatment as the omitted treatment category for three different reasons. First, the program management designated this treatment as the new program baseline. Second, all but our first hypothesis compare variations against the GAIN$^+$ baseline. Third, due to the delay in implementing the GAIN treatment, the group has the fewest observations. 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 (see Tables B.7 and B.8).

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. We thus interpret the effects of the peer treatments as conditional on the displayed savings amount, consistent with our motivation for H3.

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 at month-site level.[24] We include fixed effects for the advisors visiting the household through the vector $\mathbf{A}_i$. As two thirds of visits are conducted in pairs, the fixed effects include the first advisor ID and a dummy for whether a second advisor is present (or not). Finally, ϵ_i denotes the error term.

To explore potential mechanisms driving treatment effects, we adjust Eq. (1) by introducing indicators for combinations of treatment and heterogeneity-grouping variables. We focus on three different heterogeneity groupings: the expected financial savings from investment, the type of social benefit payment and the within-month timing when the letter is received (Carvalho et al., 2016). As an example, we define an indicator that equals 1 if the participant is in the $LOSS^+$ group and receives long-term unemployment benefits ($LOSS^+$: Unemployment), and an indicator that equals 1 if the participant is in the $LOSS^+$ group and receives some other social benefit payment ($LOSS^+$: Other Benefit). Every respondent falls into exactly one of the resulting $7 \times 2 = 14$ treatment-by-subgroup cells. We conduct the same exercise for the savings grouping (below- vs. above-median savings from replacement) and the information letter timing (visit at start vs. end of month) in separate regressions.

6. Results

We first discuss the results of the treatments that modify the information letter design in Section 6.1. Results of the reminder treatments are discussed in Section 6.2.

6.1. Information letter treatments

Table 2 shows the results of a regression of our main outcome variable of interest, refrigerator replacement, on the different information letter treatments. Fig. A.5(a) summarizes our preferred specification (6) from that table. Table B.3 and Fig. A.6 display the results of the same

[23] Fig. A.3 displays the corresponding voucher request rates.

[24] 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.

analysis for voucher request rates. Across all specifications, the GAIN$^+$ treatment is the omitted baseline group.

As displayed in our preferred specification (6) of Table 2, which includes controls for the covariates and fixed effects for a month-site interaction, all estimated treatment coefficients for refrigerator replacement display negative signs and, thus, indicate that the GAIN$^+$ baseline yields the highest refrigerator replacement rates. However, only the difference relative to the LOSS$^+$ treatment is statistically significant among the different information letter designs. At the 1-percent significance level, the LOSS$^+$ treatment decreases replacement rates by 10.7 percentage points compared to GAIN$^+$.

This statistically significant effect of the LOSS$^+$ treatment is robust across the different regression specifications in Table 2. 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 from an information letter to a replaced refrigerator by 7.6 to 12.1 percentage points. The LOSS$^+$ effect is also robust to adjustment for multiple hypothesis testing (see Table B.5).

With respect to voucher request rates, Table B.3 column (6) shows that the negative effect of the LOSS$^+$ treatment is robust for request rates and significant at the 5-percent level. The GAIN (legacy) continues to display a negative and insignificant coefficient. For the GAIN$^+$ PEER and LOSS$^+$ PEER treatments, the coefficient turns positive for voucher requests but remains insignificant. Across the different specifications, the LOSS$^+$ effect is robust, with the exception of specification (5) which includes advisor fixed effects.[25]

Table B.6 explores the linkage between request and replacement choices by displaying the results of a regression of the replacement probabilities conditional on requesting the voucher. In this table, the coefficients of the GAIN$^+$ PEER and LOSS$^+$ PEER treatments are significant and negative in our preferred regression specification. This implies that households successfully encouraged to request the voucher via the peer experience framing are more likely to ultimately fail to successfully conduct the replacement. From a program perspective, such peer experience 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.

With respect to our hypotheses, we do not find robust evidence supporting H1. While replacement rates are 2.5–5.2 percentage points higher in the management's newly introduced GAIN$^+$ frame compared to the legacy GAIN frame, this difference is not statistically significant. Second, our findings clearly oppose hypothesis H2 and the predictions of third parties: The LOSS$^+$ frame significantly reduces replacement compared to the GAIN$^+$ frame. Here, our findings thus contrast with evidence on the effectiveness of loss frames (Laibson and List, 2015) and the intuition of experts and laypeople. Third, our results fail to confirm H3. The peer experience framing in the GAIN$^+$ PEER treatment does not increase replacements as compared to the framing based on individual engineering estimates in the GAIN$^+$ baseline, while holding the communicated savings amount constant. The coefficients rather point to a negative effect but do not differ statistically from zero. Likewise, a post-estimation test on Table 2 (6) fails to reject equality of the LOSS$^+$ and LOSS$^+$ PEER treatments ($p = 0.1355$). 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 does not foster replacement rates in our sample.

These findings raise the question of why the LOSS$^+$ version of the letter backfires. Fig. A.7 plots the cumulative replacement rates under the GAIN$^+$ and LOSS$^+$ treatments after receipt of the letters. The cumulative distributions indicate a level-shift in the replacement motivation starting on the day of letter issuance. Over time, this gap between replacements in GAIN$^+$ and LOSS$^+$ further increases and then remains stable at about 10 percentage points from day 40 onward after letter issuance. The patterns suggest that the LOSS$^+$ treatment leads to an immediate shift in motivation.

We test two potential mechanisms explaining such a shift, extending our preferred regression model by introducing mutually exclusive indicators for combinations of treatment and heterogeneity-grouping variables. First, the loss formulation may be difficult to process, leading households in the LOSS$^+$ treatment to disregard the letter (and the information provided). In other words, because of its complexity, households may pay less attention to the LOSS$^+$ letter compared to the simpler GAIN$^+$ letter. If the loss formulation were indeed difficult to process, we should not expect any effects from differences in potential savings as communicated in the loss-framed letter on replacement rates.

Column (1) in Table 3 shows the regression results from combining the treatments with an indicator for having below versus above median estimated savings from replacement (see Fig. A.8 for a coefficient plot). The LOSS$^+$ treatment backfires more strongly among respondents with above median savings. The estimated coefficient for the LOSS$^+$-High Savings group shows a statistically and economically larger negative effect than that of the LOSS$^+$-Low Savings group ($p = 0.07$). Such a LOSS$^+$ intensity effect (i.e., a greater response to larger savings) suggests that households did pay attention to the treatment formulation and did not disregard the LOSS$^+$ letter. We thus rule out a lack of attention as an explanation for the backfiring of LOSS$^+$. By contrast, the finding of a LOSS$^+$ intensity effect points toward a second potential mechanism. The households may be influenced by the money at stake and the financial stress associated with "losing" money.

More specifically, a second mechanism may be poverty and scarcity-induced changes in decision-making. Consistent with arguments that financial stress can impact cognitive capacities and change time and risk attitudes (Mullainathan and Shafir, 2013; Haushofer and Fehr, 2014), Chapman et al. (2025) find that low-income households display a higher tolerance toward losses compared to households with higher incomes. Such a loss tolerance may also explain why our low-income sample reacts differently to the LOSS$^+$ treatment compared to evidence collected from samples with higher incomes.

Column (2) in Table 3 supports such a conjecture. We introduce indicators that combine treatments with a dummy variable for receiving long-term unemployment benefits, i.e., living at the minimum subsistence level, or receiving some other form of social benefit payment that is supplemental to obtained incomes or pensions (see Fig. A.9 for a coefficient plot). Notably, the LOSS$^+$ treatment shows a statistically and economically larger negative effect for recipients of long-term unemployment benefits than for recipients of supplemental benefits ($p = 0.03$) – consistent with financial scarcity explaining the backfiring result.

To further investigate scarcity as an explanation for the LOSS$^+$ backfiring, we compare refrigerator replacement rates for information letters received at the start versus at the end of the month, i.e., when the monthly benefit payment has just arrived versus when money is scarce and financial stress is high.[26] If the backfiring of the loss frame is explained by financial stress among households, we would expect a more negative effect of the loss frame at the end of a given month when liquidity concerns are usually more severe among these households.

Fig. A.10 provides descriptive evidence that replacement rates under LOSS$^+$ fall over the course of the month, while replacement rates

[25] Table B.4 shows that the LOSS$^+$ effect on request rates is further robust to applying the correction for multiple hypothesis testing by Barsbai et al. (2020), yet not robust to the correction by List et al. (2019).

[26] The payday event has been used in prior literature as exogenous variation for financial stress (Mani et al., 2020; Carvalho et al., 2016). In our setting, the relative timing of information letter receipt is quasi-random as it is induced by current wait times at local sites when households register for an audit – a metric households have limited information about and cannot influence. Our data further show that treatment assignment does not significantly predict the day-of-month of letter receipt.

Table 2
Treatment effects on refrigerator replacement.

	(1)	(2)	(3)	(4)	(5)	(6)
	Refrigerator replacement (0/1)					
GAIN$^+$	REF	REF	REF	REF	REF	REF
LOSS$^+$	−0.104**	−0.104**	−0.121***	−0.097**	−0.076*	−0.107***
	(0.042)	(0.042)	(0.040)	(0.040)	(0.042)	(0.041)
GAIN$^+$ PEER	−0.050	−0.044	−0.066	−0.054	−0.056	−0.061
	(0.043)	(0.046)	(0.046)	(0.047)	(0.048)	(0.048)
LOSS$^+$ PEER	−0.094**	−0.088*	−0.077*	−0.060	−0.069	−0.042
	(0.042)	(0.045)	(0.046)	(0.046)	(0.046)	(0.046)
GAIN (legacy)	−0.049	−0.051	−0.052	−0.030	−0.025	−0.033
	(0.049)	(0.049)	(0.048)	(0.049)	(0.050)	(0.049)
GAIN$^+$ REMINDER	−0.068**	−0.069**	−0.072**	−0.076**	−0.063*	−0.069**
	(0.034)	(0.034)	(0.034)	(0.033)	(0.033)	(0.033)
LOSS$^+$ REMINDER	−0.070	−0.070	−0.076*	−0.079*	−0.070	−0.080*
	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.043)
LOSS$^+$ PEER REMINDER	−0.041	−0.035	−0.042	−0.079*	−0.073	−0.084*
	(0.045)	(0.047)	(0.048)	(0.047)	(0.049)	(0.048)
Constant	0.242***	0.234***	0.442***	0.224***	0.221***	0.215**
	(0.031)	(0.037)	(0.099)	(0.039)	(0.043)	(0.108)
Savings Info	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Fixed Effects	No	No	No	Yes	Yes	Yes
Advisor FE	No	No	No	No	Yes	No
N	1803	1802	1761	1773	1749	1734

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. Column (2) adds a variable controlling for the communicated savings from replacement. Columns (3) and (6) 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 (4)-(6) include fixed effects of an interaction between the respective intervention site and the month when the participant is informed about his/her replacement eligibility. Column (5) additionally includes fixed effects for the main advisor visiting the household, and a dummy for whether a second advisor was present. Robust standard errors are in parentheses. Significance levels: * : $p < 0.10$, ** : $p < 0.05$, *** : $p < 0.01$.

under GAIN$^+$ remain stable over the month, and the slope is steeper for the long-term unemployed-only sample, i.e., the poorest households. Column (3) in Table 3 formalizes these patterns in regressions that combine the treatments with an indicator for receiving the letter at the start of a month versus receiving the letter at the end of a month for the full sample. Column (4) in Table 3 reports on the same regression but restricts the sample to households living on long-term unemployment benefits. In the full sample, replacement rates in the LOSS$^+$ treatment are constant between the start and end of month. However, once restricting to the long-term unemployed households, the LOSS$^+$ coefficient shows a larger negative effect at the end of the month, mirroring the downward slope in Fig. A.10 (see Fig. A.11 for a coefficient plot). While the coefficients of LOSS$^+$-Start-of-month and LOSS$^+$-End-of-month are not significantly different from each other ($p = 0.53$), we view the differential size of coefficients as suggestive evidence of aggravated financial scarcity around the payday for the poorest households in our sample.

6.2. Reminder treatments

Table 2 displays the coefficients on refrigerator replacements for the reminder treatments, and Fig. A.5(b) visualizes Specification (6). Table B.3 and Fig. A.5(b) do the same for the outcome of voucher requests. For refrigerator replacement, we observe significant and negative effects across all REMINDER treatments. Reminders in combination with the GAIN$^+$ treatment decrease energy efficiency investments by 6.9 percentage points compared to the GAIN$^+$ baseline, which is significant at the 5-percent level. The decrease in the LOSS$^+$ REMINDER and LOSS$^+$ PEER REMINDER treatments is a bit larger at respectively, 8.0 and 8.4 percentage points, all significant at the 10-percent level. The negative effect of the GAIN$^+$ REMINDER treatment is robust across the different specifications of Table 2, but does not survive multiple hypothesis correction (Table B.5). For voucher requests, the estimated effects of all reminder treatments are likewise negative but fail to reach significance in our preferred specification (Table B.3).

The findings contradict our hypothesis H4 and the predictions of third parties. The reminder decreases the effectiveness of the GAIN$^+$ treatment instead of increasing it. Furthermore, according to post-estimation tests of (6) in Table 2, the LOSS$^+$ REMINDER treatment does not significantly improve upon the LOSS$^+$ treatment ($p = 0.5151$), and the LOSS$^+$ PEER REMINDER treatment does not significantly improve upon the LOSS$^+$ PEER treatment ($p = 0.3372$).

To further investigate the unexpected findings, we split the data according to reminder type (see Tables B.7 and B.8 for voucher requests and refrigerator replacement, respectively). For refrigerator replacement, the GAIN$^+$ Letter/SMS and GAIN$^+$ Letter/SMS Tag treatments have significant negative effects at 7.2 and 9.1 percentage points relative to GAIN$^+$, while the GAIN$^+$ Tag treatment is insignificant and closer to 0. As tested above, both LOSS$^+$ treatments combined with a tag do not significantly affect replacement rates compared to the simple LOSS$^+$ effect.[27] In summary, none of the tag reminders significantly changes replacement rates in either direction in comparison to their respective information letter-only version. Exclusively the Letter/SMS reminders drive the significant negative findings and cause the rejection of H4.

A potential explanation for such an empirical finding may be found in negative anticipation effects as described in Section 3.3 and proposed by Ericson (2017). The participants in the SMS/Letter treatment anticipated receiving a reminder about the replacement voucher as announced

[27] We do not observe a combination of the LOSS$^+$ and Letter/SMS Reminder.

Table 3
Potential mechanisms of treatment effects.

	(1)	(2)	(3)	(4)
	Refrigerator replacement (0/1)			
GAIN$^+$: Low Savings	REF			
GAIN$^+$: High Savings	−0.008			
	(0.067)			
LOSS$^+$: Low Savings	−0.041			
	(0.059)			
LOSS$^+$: High Savings	−0.149**			
	(0.061)			
GAIN$^+$: Other Benefit		REF		
GAIN$^+$: Unemployment		−0.047		
		(0.068)		
LOSS$^+$: Other Benefit		−0.041		
		(0.075)		
LOSS$^+$: Unemployment		−0.186***		
		(0.065)		
GAIN$^+$: Start of Month			REF	REF
GAIN$^+$: End of Month			0.012	0.011
			(0.060)	(0.078)
LOSS$^+$: Start of Month			−0.102*	−0.085
			(0.059)	(0.074)
LOSS$^+$: End of Month			−0.100*	−0.123**
			(0.055)	(0.059)
Constant	0.217*	0.277**	0.193*	0.202
	(0.125)	(0.114)	(0.113)	(0.151)
Savings Info	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Advisor FE	No	No	No	No
Sample	w/o PEER	All	All	Unemployed
N	1244	1734	1734	1092

Note: Linear probability models of refrigerator replacement (yes/no) on treatments. Treatments are included as mutually exclusive indicators for the respective treatment group and (1) above- vs. below-median savings from replacement, (2) long-term unemployment benefits vs. other types of social benefit payments, (3)-(4) letter receipt on days 1–15 vs. day 16–31 of a month. Due to design differences in communicated savings, (1) omits the PEER treatments. Column (4) is estimated on the sample of long-term unemployed only. All columns control for the communicated savings from replacement, 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 and include fixed effects of an interaction between the respective intervention site and the month when the participant is informed about his/her replacement eligibility. Robust standard errors are in parentheses. Significance levels: * : $p < 0.10$, ** : $p < 0.05$, *** : $p < 0.01$.

during the second visit, and thus postponed the decision to request the voucher, allocated less mental effort to remember, and, ultimately, never followed through with the investment choice. A comparison of the timing of voucher requests after the second visit reveals that, directly from the start, cumulative voucher request rates in the GAIN$^+$ Letter/SMS reminder treatment are consistently lower than in the simple GAIN$^+$ treatment, and the differential further grows during the first month (see Fig. A.12). Interestingly, no catch-up is visible during and after the period when the reminders were sent out. In comparison, the tag reminder closely follows the GAIN$^+$ baseline. As the tag was present in the household from the start, no negative anticipation effect arose. In summary, the results are consistent with a negative anticipation effect of the Letter/SMS reminder. The announcement may have caused households to anticipate the reminder, postpone requesting the voucher and making the replacement choice and reduce the mental capacity spent on thinking about the replacement choice. Ultimately, this provides one explanation for the backfiring effect we observe in our data.

7. Ranking design alternatives

In line with a recent literature (Hahn et al., 2026), we use the marginal value of public funds (MVPF) to rank the different design alternatives for the RRP. We proceed in two steps. The first is to arrive at plausible MVPF estimates for the baseline treatment (GAIN$^+$) of the RCT. The detailed calculations giving rise to these estimates are relegated to Appendix C. The second step uses the treatment effects derived in Section 6 and additional program information to rank the MVPFs of the design alternatives relative to the baseline and to each other.

Table 4 reports on the results of the two steps. We first focus, as a benchmark, on the MVPF estimate of the baseline treatment, which relies on our preferred way of estimating the MVPF. This uses, for reasons of comparability with Hahn et al. (2026), the midpoint of the price elasticity of demand and the US EPA Social Cost of Carbon estimate for 2024 under a 2.0 percent near-term discount rate (US$208, EPA (2023)), and a mean price elasticity of −1.66 drawn from Chlond et al. (2025).[28]

Compared to previous MVPF estimates for appliance replacement programs, our benchmark MVPF comes in at a low level, at 0.63. MVPF calculations for refrigerator replacement programs have in the past returned values between 0.958 for the California Energy Savings Assistance Program for Refrigerators (Blonz, 2023) and 1.042 for refrigerators replaced under the Cash for Appliances program (Houde and Aldy, 2017). Typically, refrigerator replacement programs perform below the midpoint of replacement programs for appliances generally (1.164). The baseline MVPF of the RRP therefore falls substantially below that of other such programs. Detailed calculations highlight the role of substantial encouragement efforts, that is, marketing costs, of the RRP in explaining this outcome. Much of the public funds is spent on marketing to eligible households that ultimately do not replace their appliances and therefore do not collect the subsidy.

At the same time, the RRP is specifically designed as a social assistance program. A complete welfare assessment therefore has to include appropriate social welfare weights for the target group that benefits from the program. Recent estimates of the welfare weights implicit in German fiscal policy (Isaak and Jessen, 2024) put these weights at around 2.9 (1.4) for the first (second) income decile. These deciles make up the main target population of the RRP. Including their welfare weights could therefore provide minimum fiscal justification for the program by lifting welfare-weighted MVPF across the unit threshold.

Comparing the MVPF of the baseline with those of its design alternatives reported in Table 4 shows that the latter generally perform worse in MVPF terms. Behind this finding are the two factors that determine the replacement-cost ratio of each treatment, i.e., the cost per percentage point of replacement rate. Policies that achieve a higher replacement rate at the same "marketing" cost or the same replacement rate at lower marketing cost achieve a higher MVPF. To rank better, costlier design alternatives, viz. those including reminders, need to sufficiently outperform their simpler variants. As the ratios show, however, this is not the case. Simple variations on the information letter without reminder tend to perform as well as more elaborate designs with reminders.

The last column of Table 4 reports the relative rank of each design alternative. This ranking reflects the patterns of replacement rates and marketing costs as the main drivers of relative performance. The baseline treatment GAIN $^+$ dominates the ranking, followed by two other GAIN frame treatments. The five best performing design alternatives do not involve a follow-up with households. Design alternatives with reminders, on the other hand, are dragged down – in terms of MVPF and ranking – by their less favorable replacement-cost ratio.

[28] Table C.1 in Appendix C shows the considerable range of plausible MVPF estimates for the baseline RRP beyond our benchmark estimate. This range has three main drivers: (i) the 95-percent confidence interval of the price elasticity of refrigerator replacement demand covers a wide range; (ii) there can be disagreement about the in- or exclusion of the rebound effect of reduced CO_2 emissions in the residential sector; and if rebound effects are not considered, (iii) there is a wide range of estimates of the social cost of carbon (SCC), including that of the EPA (2023) embodied in our benchmark.

Table 4
Treatment replacement rate, marketing cost, replacement-cost ratio, MVPF, and rank.

	Replacement rate	Cost	Ratio	MVPF	Rank
GAIN$^+$ (baseline)	0.24	20	1.2	0.63	1
LOSS$^+$	0.14	20	0.7	0.49	5
GAIN$^+$ PEER	0.19	20	0.95	0.57	2
LOSS$^+$ PEER	0.15	20	0.75	0.51	4
GAIN (legacy)	0.19	20	0.95	0.57	2
GAIN$^+$ REMINDER	0.17	30	0.57	0.44	7
LOSS$^+$ REMINDER	0.17	30	0.57	0.44	7
LOSS$^+$ PEER REMINDER	0.20	30	0.67	0.49	6

Note: Replacement rates are the empirically observed shares of eligible households that replace their refrigerator under the RRP design alternative, as reported in Fig. A.4. Costs are internal estimates of the marketing cost or cost of encouragement of each design alternative (in euros). The ratio is the replacement rate in percent divided by cost. MVPF is the preferred estimate based on the EPA (2023) Social Cost of Carbon estimate for 2024 under a 2.0 near-term discount rate and mean price elasticity. See Appendix C for detailed calculations. Rank is the relative position of the ratio, ordered from greater to smaller, across the design alternatives.

8. Conclusion

When trying to design or improve programs targeting specific groups in society, such as economically disadvantaged households, it can be tempting for program managers to deploy behavioral insights without additional guidance. Their attraction is understandable: they have repeatedly proven successful among the general population, tend to be low-cost, and appear intuitively suited to many program contexts. Yet, their unguided deployment can lead to undesirable outcomes: they may fail to generate the intended impacts among the target group, either in terms of magnitude or even in terms of direction or may turn out to be an inefficient use of public funds. When the target group is vulnerable or exposed, a failure of external validity carries additional ethical significance.

Through the implementation and evaluation of the co-designed RCT presented in this paper, policy-makers were alerted to problems with effectiveness and efficiency of popular behavioral insights before such insights were deployed at scale among the target population. Partnering with the management of one of the world's largest energy efficiency assistance programs, we engaged with policy-makers at a point when the program was about to undergo a design overhaul. Observing management's stringent budgetary, practical, political and ethical requirements, the resulting co-design process built on the rich literature on behavioral insights, on historical program evidence, on practitioner knowledge, and on expert and layperson predictions. The outcome of the process was an RCT testing eight behaviorally informed program alternatives that had been pre-screened for compliance with the SANS conditions. Their evaluation of effectiveness and efficiency then provided the basis for the possibility of guided deployment. In sum, this paper illustrates how applying established and recent techniques from the literature can help the policy-maker detect problems of external validity and avoid the dangers of unguided deployment.

The evidence on effectiveness generated by the resulting RCT illustrates that the external validity of even pre-screened behavioral insights can be under threat. Among the program's target population of low-income households, loss framing, an insight with high expected validity, significantly backfires in comparison to the management baseline. Further analyses suggest that financial stress and scarcity impacting loss tolerance as well as time and risk attitudes may be one of the mechanisms of such backfiring. Reminders – another mainstay of behavioral insights – likewise decrease program effectiveness. A likely reason that is consistent with our data is that the announcement of reminders reduces own investment in imperfect memory. Beyond questions of effectiveness, the tested program alternatives also threaten the efficient use of public funds dedicated to the program as measured through their marginal value of public funds (MVPF). This is the case as the program alternatives both fail to deliver improvements in effectiveness and fail to compensate for higher program costs. This is most evident in the low MVPF of reminders which reflects not just poor effectiveness but also the additional cost of sending reminders.

Taken together, the RCT findings highlight the risks that unguided deployment of behavioral insights incurs in programs such as the RRP, not least because of the ethically and politically sensitive nature of its target population. Had loss framing or reminders been adopted directly on the strength of their track record elsewhere – the path a conventional behavioral-insights engagement might have taken – the program would have deployed alternatives that reduced take-up among a population it was designed to support, without the ex-post evidence needed to detect this before scale-up. The guided approach we implemented here – pre-screening for scalability, piloting through an RCT, and evaluating both effectiveness and MVPF before deployment – allowed these problems to surface in time. This procedural finding likely contains a lesson for policy practice: for programs in new settings a guided approach based on pre-screening, piloting and ex-post evaluation may well be worth considering in order to guard against unexpected back-firing. This appears to be particularly valuable when targeting under-represented or vulnerable groups. We observe a divergence in expected performance across two mechanistically distinct insights in a population with well-documented exposure to financial stress and scarcity and the observed behavioral patterns are consistent with the broader concern in the literature that evidence bases built on general populations may not transfer cleanly to specific, under-represented target groups (Bryan et al., 2021; Hallsworth, 2023). Future investigations of the loss frame effect could investigate to what extent it could be mitigated by different phrasing.

Declaration of competing interest

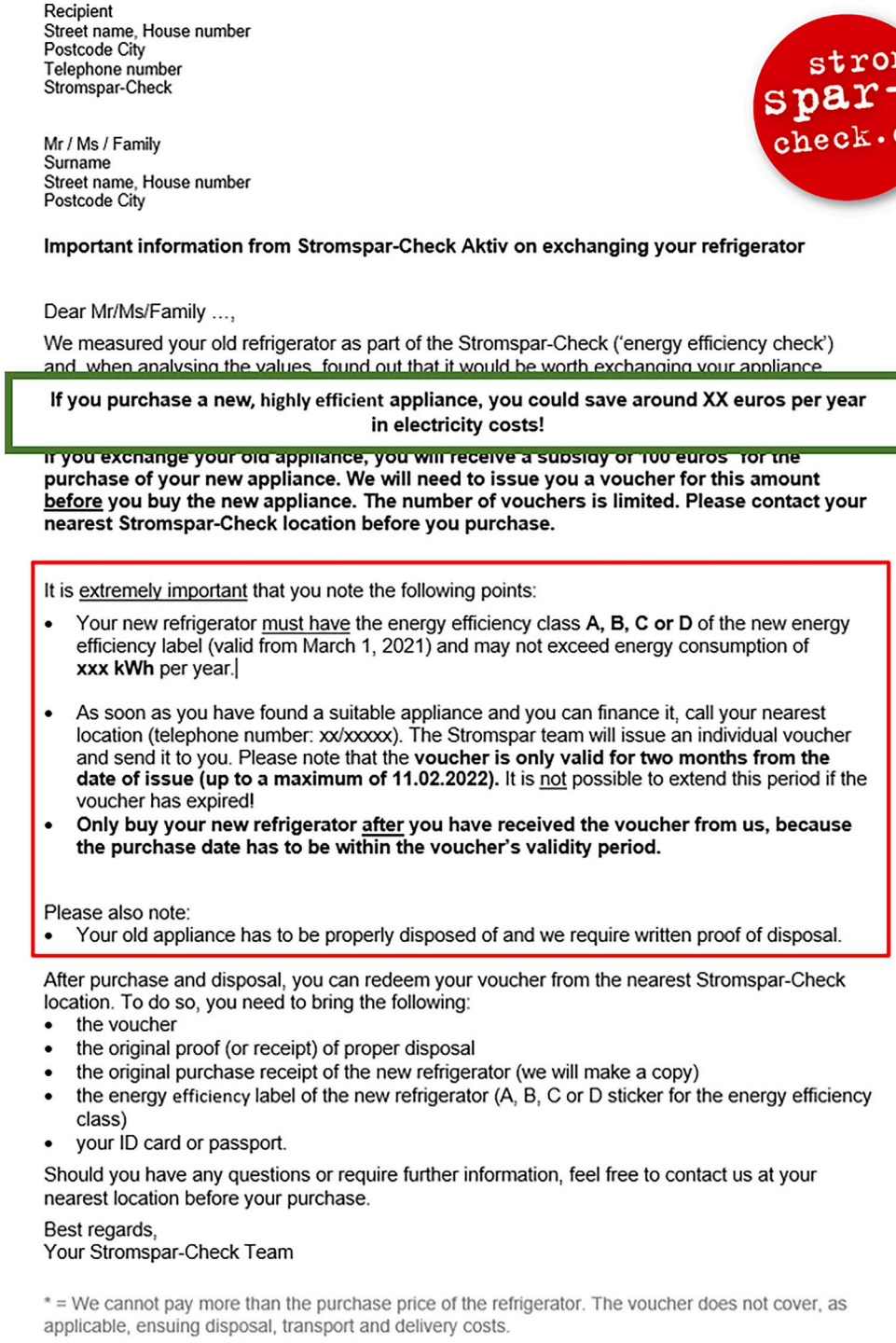
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank the SSC management and staff for the partnership. We thank John A. List, Rob. D. Metcalfe, Michael K. Price, Daan van Soest, Alessandro Tavoni and two anonymous reviewers for their valuable comments. We thank participants at EAERE 2023, AFE 2023 and 2024, NOeG 2024, and seminar participants at Amsterdam, Bochum, Bologna, Grenoble, Heidelberg, Kiel, Mannheim, Würzburg and Zurich for valuable feedback. Max Stellbrink and Carina Haller provided excellent research assistance.

Appendix A

A.1. Figures



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

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 before you buy the new appliance. The number of vouchers is limited. Please contact your nearest Stromspar-Check location before you purchase.

It is extremely important that you note the following points:

- Your new refrigerator must have 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 not possible to extend this period if the voucher has expired!
- **Only buy your new refrigerator after 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.

Fig. A.1. Letter informing about voucher eligibility. Note: Green highlights are added to the original letter. Red highlights are not added and are part of the original letter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

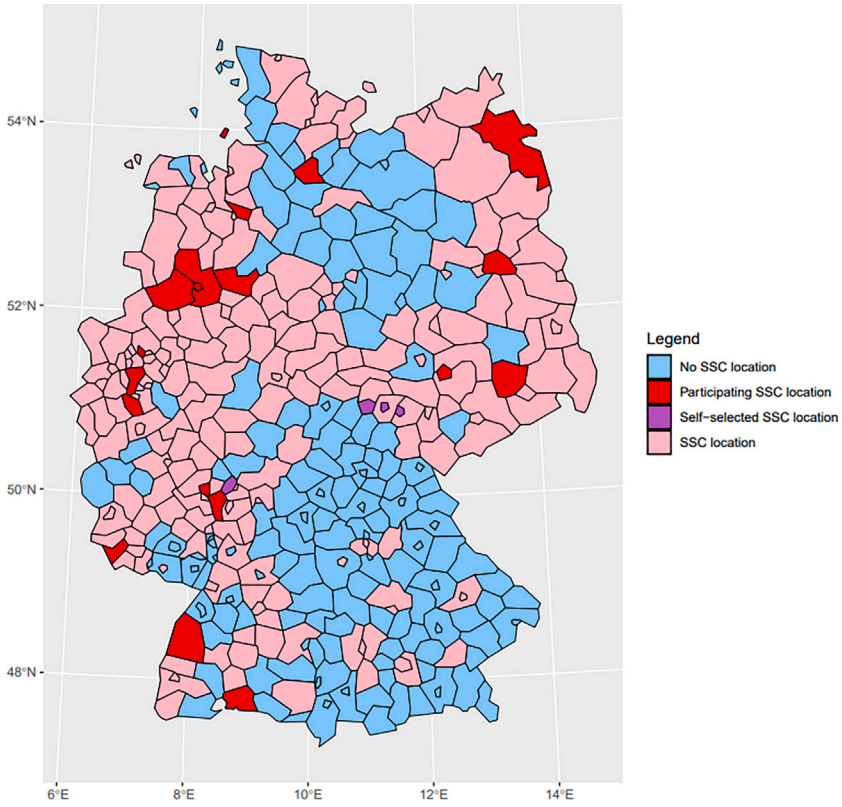
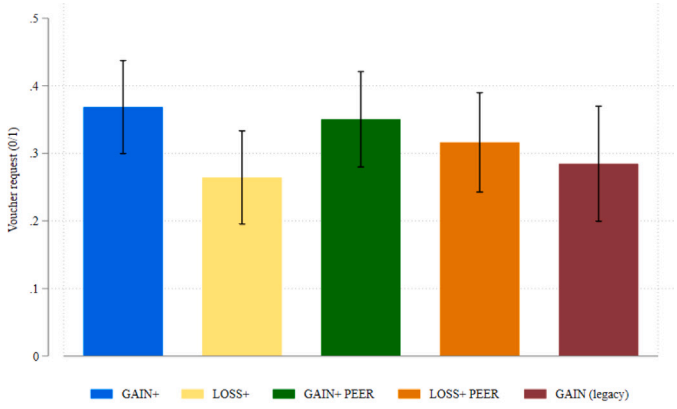


Fig. A.2. Program sites of the "Energy-Saving-Check".

(a) Information letter treatments.



(b) Reminder treatments.

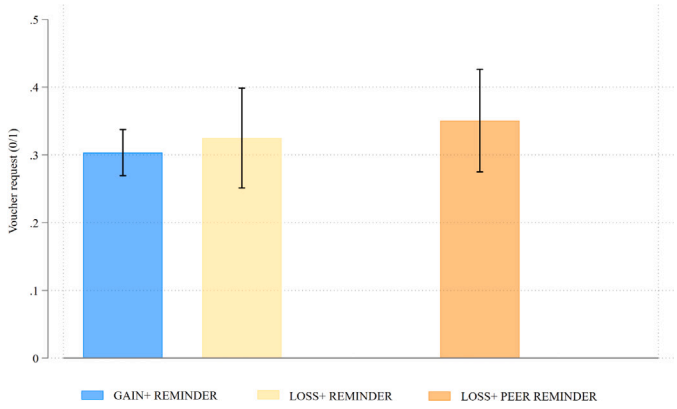
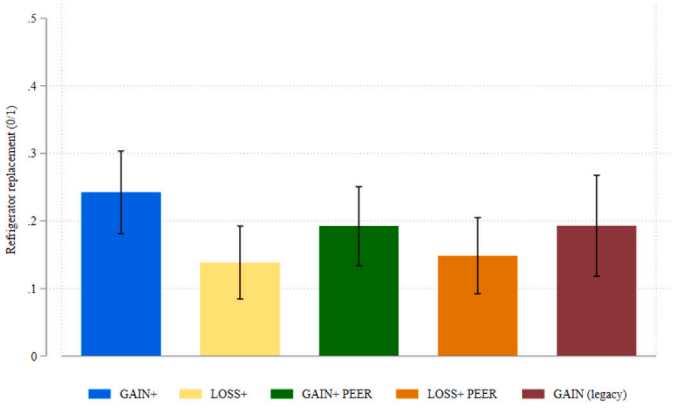


Fig. A.3. Voucher request rates by treatment group.

(a) Information letter treatments.



(b) Reminder treatments.

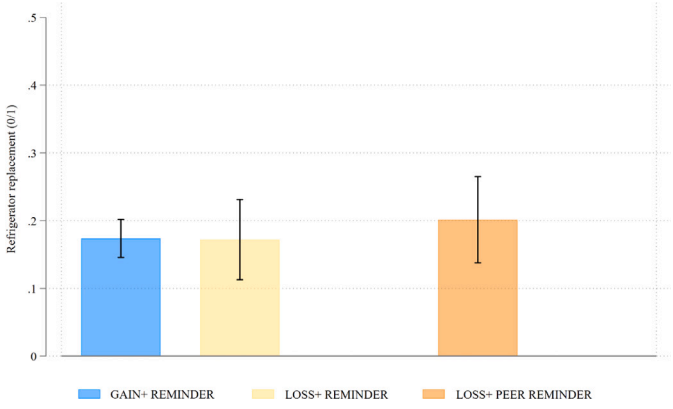
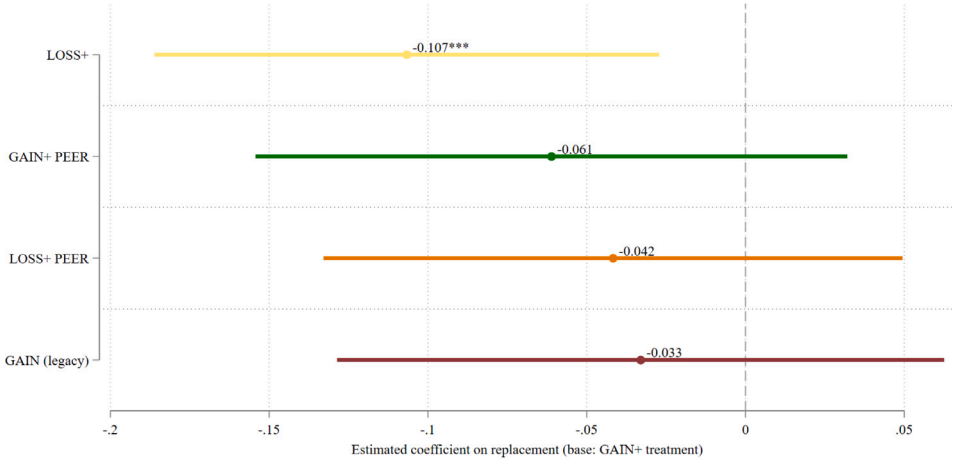


Fig. A.4. Refrigerator replacement rates by treatment group.

(a) Information letter treatments.



(b) Reminder treatments.

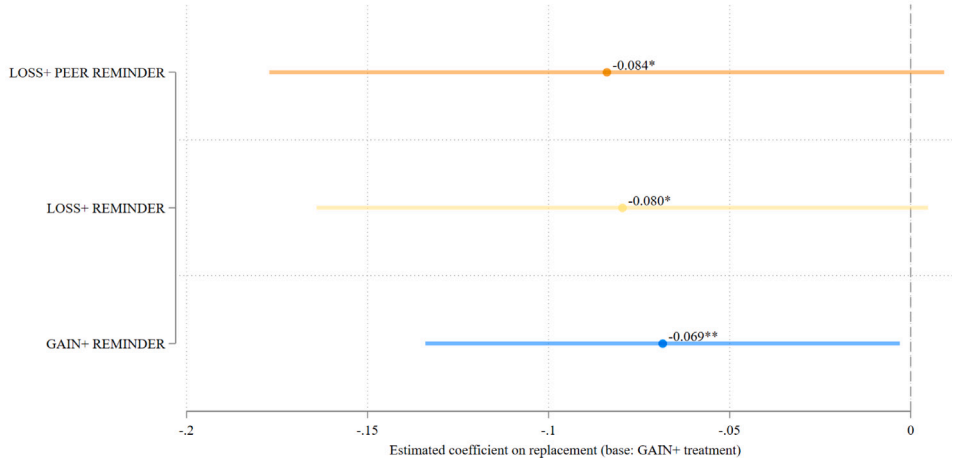
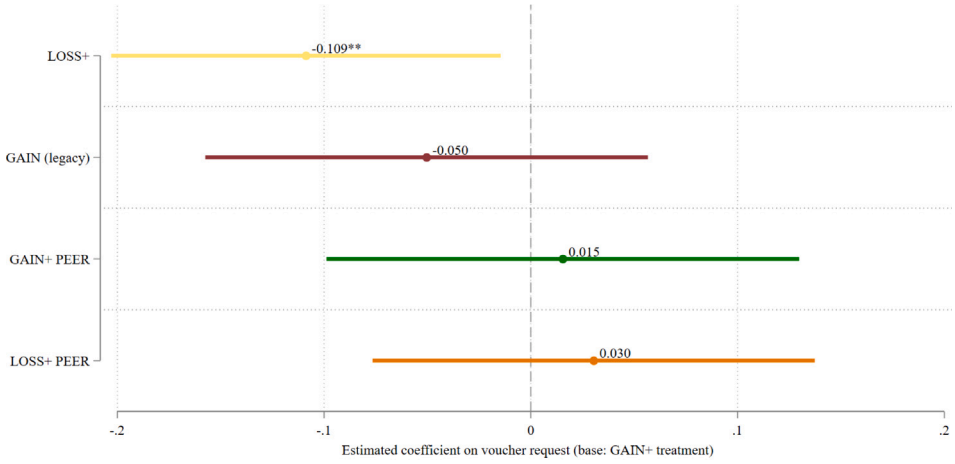


Fig. A.5. Coefficient plot of refrigerator replacements by treatment. Note: Displayed are the estimated coefficients from column (6) of Table 2. The coefficients are sorted by effect size.

(a) Information letter treatments.



(b) Reminder treatments.

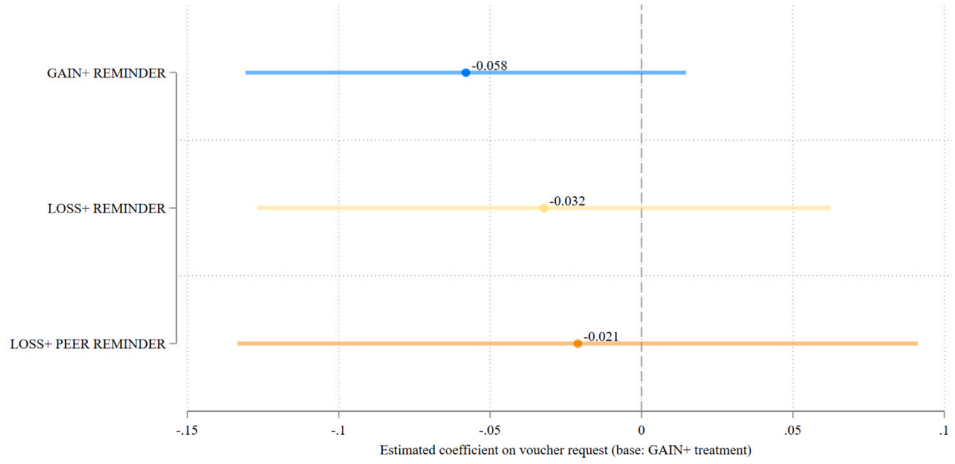


Fig. A.6. Coefficient plot of voucher requests by treatment. Note: Displayed are the estimated coefficients from column (6) of Table B.3. The coefficients are sorted by effect size.

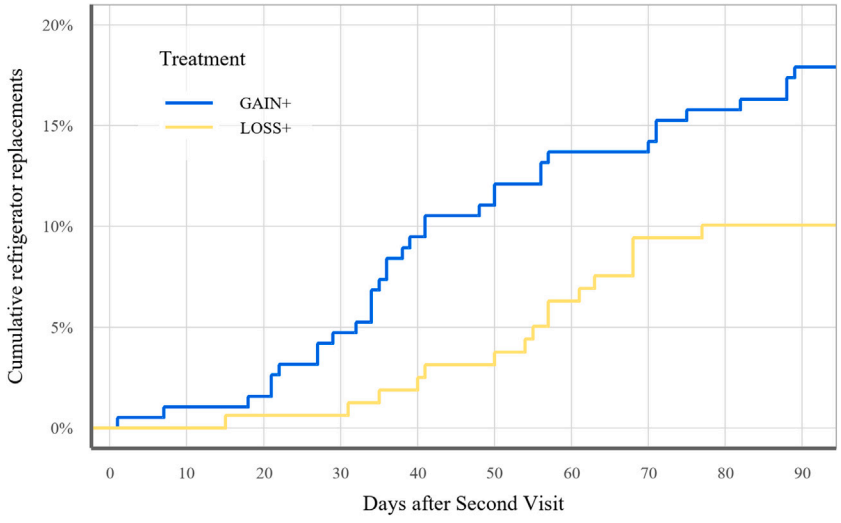


Fig. A.7. Cumulative replacement rates against days after the second visit (GAIN + vs. LOSS +). Note: Displayed are the cumulative replacement rates as a function of days after the second visit for treatments GAIN + and LOSS +.

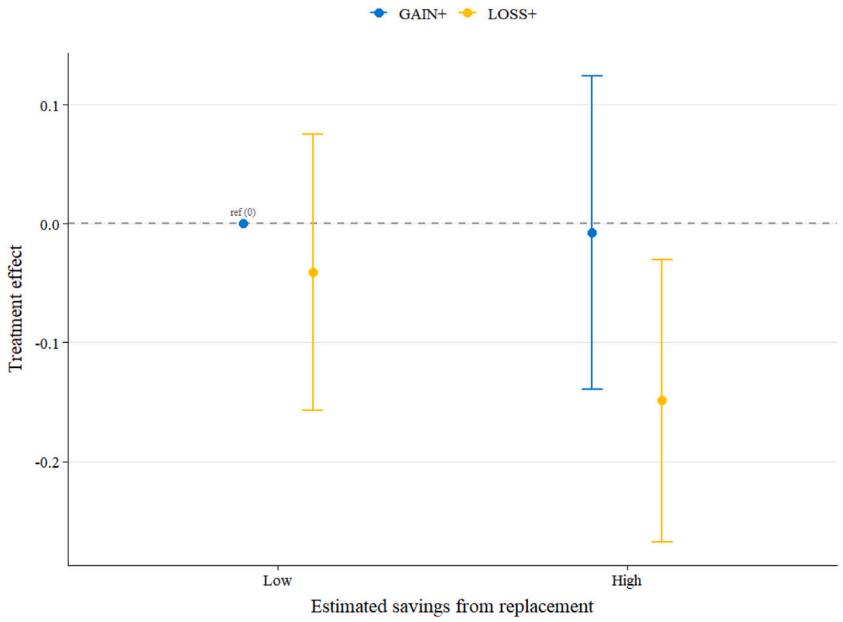


Fig. A.8. Coefficient plot of heterogeneous treatment effects on refrigerator replacement rates by estimated savings from replacement. Note: This figure plots coefficients for cross-groups of $LOSS^+$ and $GAIN^+$ with above- and below-median estimated savings from replacement.

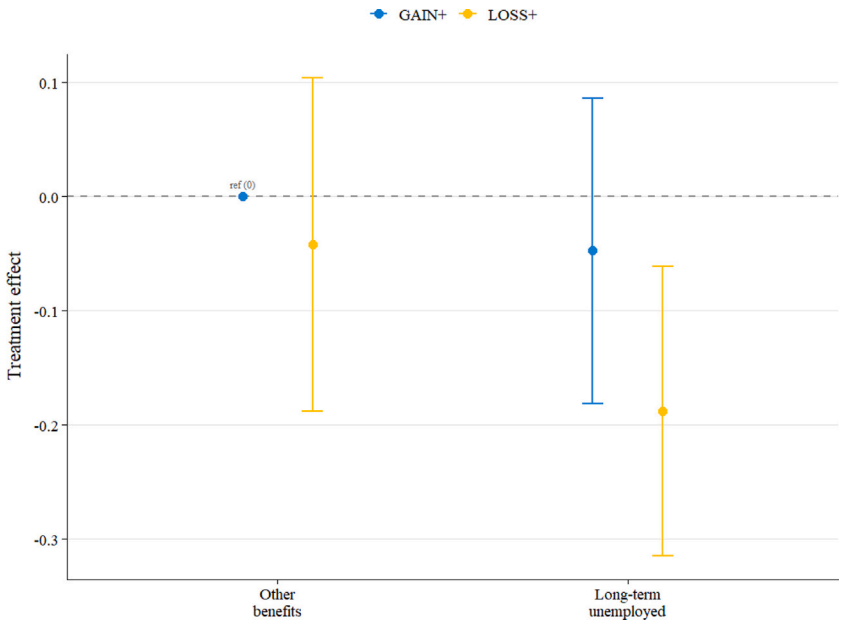
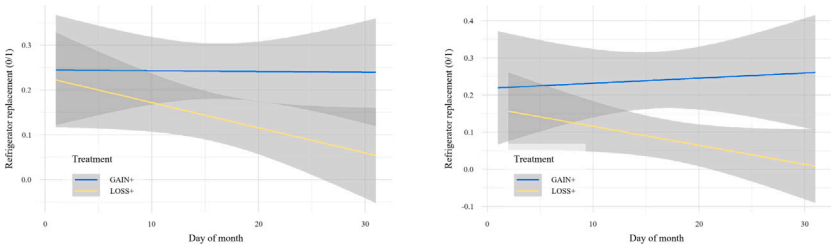


Fig. A.9. Coefficient plot of heterogeneous treatment effects on refrigerator replacement rates by transfer type. Note: This figure plots coefficients for cross-groups of $LOSS^+$ and $GAIN^+$ among recipients of long-term unemployment benefits and other benefits.



(a) Full sample (b) Recipients of long-term unemployment benefits

Fig. A.10. Refrigerator replacement rates by day-of-month. Note: Displayed are the linear fitted lines through the raw data points for the outcome refrigerator replacement by day of month of treatment receipt. The left panel uses the full sample, the right panel uses only recipient households of long-term unemployment benefits. 95-percent confidence intervals are given in gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

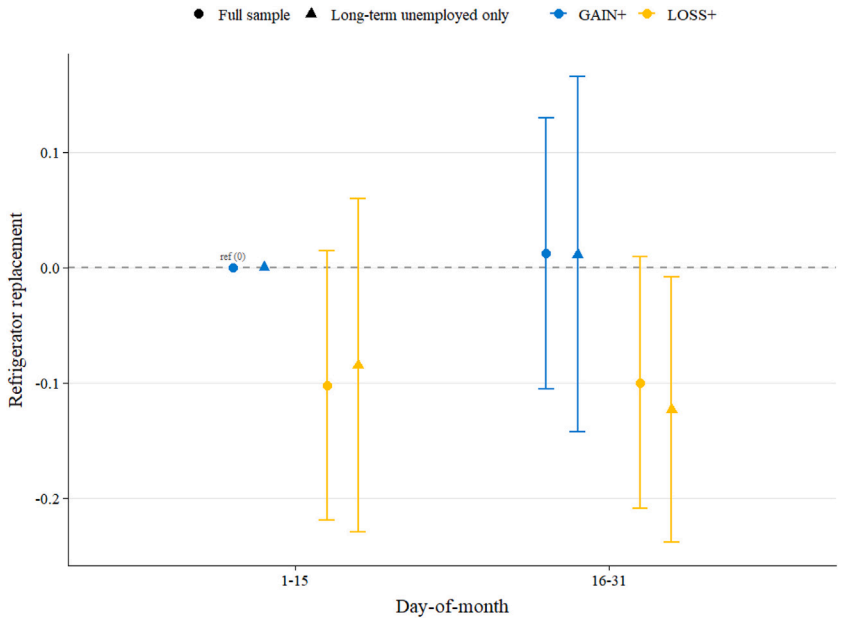


Fig. A.11. Coefficient plot of heterogeneous treatment effects on refrigerator replacement rates by timing of receipt of the information letter. Note: This figure plots coefficients for cross-groups of LOSS$^+$ and GAIN$^+$ with start-of-month and end-of-month receipt of the information letter.

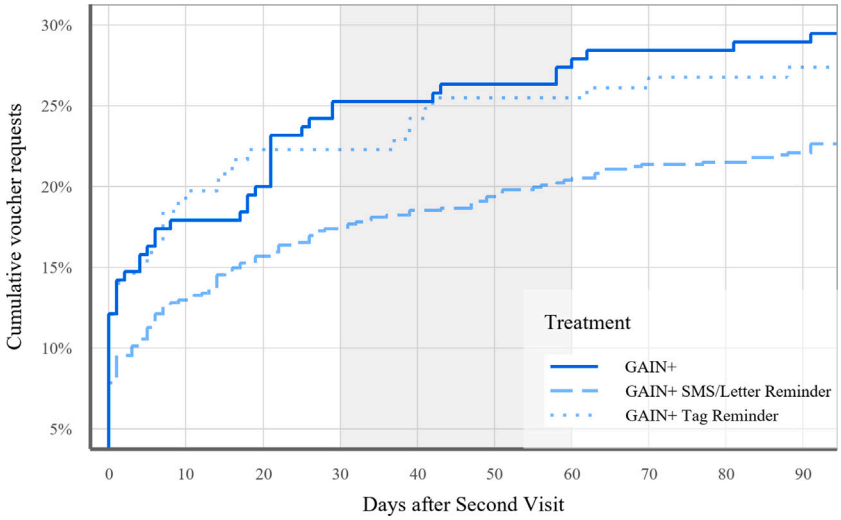


Fig. A.12. Cumulative voucher request rates against days after the second visit (GAIN + vs. GAIN + reminders). Note: Displayed are the cumulative voucher request rates as a function of days after the second visit for treatments GAIN$^+$, GAIN$^+$ Letter/SMS reminders, and GAIN$^+$ Tag reminders. The gray shaded area marks the period in which households receive the Letter/SMS reminders. Tag reminders are installed in treated households' refrigerators on the second visit, and therefore are present from day 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix B. Tables

Table B.1
Timeline.

Time Period	Key Activities
April 2020	• Co-design process with program officials and local site managers
July–December 2021	• Series of follow-up workshops to co-design process • Pre-test pilot at Frankfurt local site • Randomized selection of sites
December 2021	• Invitation to participate for 30 randomly drawn sites • 23 out of the 30 sites accepted the invitation. • Preparation of Q&A sessions and training materials for site managers • Implementation of Q&A sessions and training sessions to familiarize site managers with interventions
January 2022	• Implementation of Q&A sessions and training sessions to familiarize site managers with interventions • Programming of database • Shipping of treatment materials • Exclusion of two sites (Munich due to parallel testing; Groß-Gerau-Kreis due to lack of site management).
January–March 2022	• Training period for site managers to inform, organize, and train their local staff.
April 1, 2022–February 15, 2023	• Implementation of field experiment • Local sites stopped voucher issuance on February 15, 2023

Table B.2
Multiple hypothesis corrected 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
	(0.134)	(0.108)	(0.909)	(0.416)	(0.665)	(0.959)	(0.446)	(0.675)
Energy consumption (kWh)	–52.260	219.960	–60.091	–126.032	–84.370	77.374	–92.195	–79.227
	(0.775)	(0.479)	(0.988)	(0.759)	(0.748)	(0.817)	(0.959)	(0.940)
Electric water heating: (1 = yes)	0.049	–0.022	0.025	0.056	–0.021	–0.030	–0.032	0.046
	(0.684)	(0.908)	(0.982)	(0.542)	(0.604)	(0.703)	(0.932)	(0.809)
Estimated savings (kWh)	37.361*	–0.413	–0.397	0.310	16.051	–18.097*	1.631	10.618
	(0.092)	(0.973)	(1)	(0.982)	(0.783)	(0.070)	(0.988)	(0.896)
No. persons (count)	–0.157	0.151	–0.001	–0.101	0.127	–0.009	0.017	–0.072
	(0.550)	(0.716)	(0.995)	(0.882)	(0.800)	(0.920)	(0.911)	(0.952)
Living space (m^2)	–2.150	2.861	0.768	–3.001	–3.551	1.723	–0.925	–0.009
	(0.737)	(0.590)	(0.974)	(0.535)	(0.364)	(0.641)	(0.951)	(0.997)
Transfer scheme (1–7)	0.382	–0.177	0.182	–0.129	–0.244	0.036	0.132	–0.101
	(0.725)	(0.863)	(0.981)	(0.851)	(0.791)	(0.980)	(0.967)	(0.935)
Unemployment benefits (1 = yes)	–0.052	0.014	–0.018	0.023	0.049	–0.005	–0.030	0.011
	(0.768)	(0.914)	(0.980)	(0.819)	(0.658)	(0.965)	(0.913)	(0.955)
N per group	109	190	159	177	155	702	157	154
N total	1803	1803	1803	1803	1803	1803	1803	1803

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, transfer type category and an indicator for receiving unemployment benefits by treatment group compared to all other groups. p-values that are adjusted for multiple hypothesis testing following List et al. (2019) are in parentheses. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$. Before adjusting for multiple hypothesis testing, differences in energy prices are significant for GAIN, GAIN$^+$, and GAIN$^+$ PEER, as well as estimated savings for GAIN and GAIN$^+$ REMINDER, and energy consumption for GAIN$^+$.

Table B.3
Treatment effects on voucher request.

	(1)	(2)	(3)	(4)	(5)	(6)
	Voucher request (0/1)					
GAIN$^+$	REF	REF	REF	REF	REF	REF
LOSS$^+$	−0.104**	−0.105**	−0.126***	−0.103**	−0.048	−0.109**
	(0.050)	(0.050)	(0.047)	(0.048)	(0.047)	(0.048)
GAIN$^+$ PEER	−0.018	0.031	0.012	0.023	0.028	0.015
	(0.050)	(0.053)	(0.054)	(0.057)	(0.058)	(0.058)
LOSS$^+$ PEER	−0.052	−0.004	0.004	0.021	0.007	0.030
	(0.051)	(0.054)	(0.054)	(0.054)	(0.052)	(0.055)
GAIN (legacy)	−0.084	−0.098*	−0.086	−0.048	−0.014	−0.050
	(0.056)	(0.056)	(0.054)	(0.055)	(0.053)	(0.055)
GAIN$^+$ REMINDER	−0.065*	−0.065*	−0.064*	−0.066*	−0.024	−0.058
	(0.039)	(0.039)	(0.038)	(0.037)	(0.036)	(0.037)
LOSS$^+$ REMINDER	−0.044	−0.044	−0.055	−0.031	−0.004	−0.032
	(0.051)	(0.051)	(0.049)	(0.048)	(0.047)	(0.048)
LOSS$^+$ PEER REMINDER	−0.018	0.031	0.029	−0.016	−0.003	−0.021
	(0.052)	(0.055)	(0.055)	(0.056)	(0.057)	(0.057)
Constant	0.368***	0.301***	0.739***	0.293***	0.285***	0.430***
	(0.035)	(0.043)	(0.111)	(0.046)	(0.051)	(0.115)
Savings Info	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Fixed Effects	No	No	No	Yes	Yes	Yes
Advisor FE	No	No	No	No	Yes	No
N	1803	1802	1761	1773	1749	1734

Note: Linear probability models of voucher request (yes/no) on treatments. The treatments are included as indicators for the respective treatment groups. The GAIN$^+$ treatment is the omitted reference treatment group. Column (2) adds a variable controlling for the communicated savings from replacement. Columns (3) and (6) 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 (4)-(6) include fixed effects for an interaction between the respective intervention sites and the month when the participant is informed about his/her replacement eligibility. Column (5) additionally includes fixed effects for the main advisor visiting the household, and a dummy for whether a second advisor was present. Robust standard errors are in parentheses. Significance levels: $^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

Table B.4
Multiple hypothesis corrected p-values for voucher request.

	(1) Unadj. p-value	(2) p-value following List et al. (2019)	(3) p-value following Barsbai et al. (2020)
LOSS$^+$	0.024	0.169	0.065
GAIN$^+$ PEER	0.791	0.911	0.994
LOSS$^+$ PEER	0.577	0.716	0.953
GAIN (legacy)	0.356	0.468	0.388
GAIN$^+$ REMINDER	0.118	0.377	0.311
LOSS$^+$ REMINDER	0.504	0.746	0.629
LOSS$^+$ PEER REMINDER	0.713	0.734	0.954

Note: Displayed are p-values of the treatment effects on voucher request. Column (1) displays the unadjusted p-values of the regression results in column (6) of Table B.3. Column (2) provides an adjustment of the results in column (1) of Table B.3 based on Theorem 3.1 in List et al. (2019) and using Stata package mhtexp. Column (3) provides an adjustment of the results in column (6) of Table B.3 based on Theorem 3.1 in List et al. (2019), the expansion provided by Barsbai et al. (2020) and using the Stata package mhtreg.

Table B.5
Multiple hypothesis corrected p-values for refrigerator replacement.

	(1) Unadj. p-value	(2) p-value following List et al. (2019)	(3) p-value following Barsbai et al. (2020)
LOSS$^+$	0.009	0.058	0.016
GAIN$^+$ PEER	0.198	0.504	0.251
LOSS$^+$ PEER	0.370	0.134	0.240
GAIN (legacy)	0.498	0.498	0.499
GAIN$^+$ REMINDER	0.040	0.156	0.153
LOSS$^+$ REMINDER	0.064	0.307	0.248
LOSS$^+$ PEER REMINDER	0.077	0.351	0.335

Note: Displayed are p-values of the treatment effects on refrigerator replacement. Column (1) displays the unadjusted p-values of the regression results in column (6) of Table 2. Column (2) provides an adjustment of the results in column (1) of Table 2 based on Theorem 3.1 in List et al. (2019) and using Stata package mhtexp. Column (3) provides an adjustment of the results in column (6) of Table 2 based on Theorem 3.1 in List et al. (2019), the expansion provided by Barsbai et al. (2020) and using the Stata package mhtreg.

Table B.6
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	−0.133	−0.165	−0.210*
	(0.095)	(0.096)	(0.115)	(0.118)
GAIN$^+$ PEER	−0.179*	−0.232**	−0.191	−0.267*
	(0.091)	(0.099)	(0.120)	(0.137)
LOSS$^+$ PEER	−0.257***	−0.233**	−0.268**	−0.231*
	(0.097)	(0.104)	(0.126)	(0.133)
GAIN (legacy)	0.043	0.012	−0.140	−0.163
	(0.100)	(0.100)	(0.126)	(0.127)
GAIN$^+$ REMINDER	−0.084	−0.105	−0.147*	−0.142*
	(0.066)	(0.069)	(0.079)	(0.083)
LOSS$^+$ REMINDER	−0.122	−0.121	−0.153	−0.150
	(0.090)	(0.092)	(0.097)	(0.101)
LOSS$^+$ PEER REMINDER	−0.154	−0.189*	−0.239**	−0.265**
	(0.094)	(0.102)	(0.116)	(0.125)
Constant	0.754***	0.562***	0.760***	0.466*
	(0.072)	(0.196)	(0.093)	(0.279)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
Advisor FE	No	No	No	No
Request = 1	Yes	Yes	Yes	Yes
N	571	555	519	503

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 an interaction between the respective intervention site and the month when the participant is informed about his/her replacement eligibility. Robust standard errors are in parentheses. Significance levels: * : $p < 0.10$, ** : $p < 0.05$, *** : $p < 0.01$.

Table B.7
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**	−0.126***	−0.103**	−0.109**
	(0.050)	(0.047)	(0.048)	(0.048)
GAIN$^+$ PEER	0.031	0.012	0.023	0.015
	(0.053)	(0.055)	(0.057)	(0.058)
LOSS$^+$ PEER	−0.004	0.004	0.020	0.029
	(0.054)	(0.054)	(0.054)	(0.055)
GAIN (legacy)	−0.098*	−0.086	−0.049	−0.051
	(0.056)	(0.054)	(0.055)	(0.055)
GAIN$^+$ Letter/SMS	−0.049	−0.045	−0.045	−0.034
	(0.043)	(0.042)	(0.042)	(0.042)
GAIN$^+$ Tag	−0.061	−0.063	−0.078*	−0.064
	(0.049)	(0.048)	(0.047)	(0.047)
GAIN$^+$ Letter/SMS Tag	−0.099**	−0.099**	−0.092**	−0.095**
	(0.048)	(0.047)	(0.047)	(0.047)
LOSS$^+$ Tag	−0.044	−0.054	−0.031	−0.032
	(0.051)	(0.049)	(0.048)	(0.048)
LOSS$^+$ PEER Tag	0.031	0.030	−0.017	−0.022
	(0.055)	(0.055)	(0.056)	(0.057)
Constant	0.301***	0.742***	0.294***	0.431***
	(0.043)	(0.112)	(0.046)	(0.115)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
Advisor FE	No	No	No	No
N	1802	1761	1773	1734

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 an interaction between the respective intervention site and the month when the participant is informed about his/her replacement eligibility. Robust standard errors are in parentheses. Significance levels: * : $p < 0.10$, ** : $p < 0.05$, *** : $p < 0.01$.

Table B.8
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**	−0.121***	−0.097**	−0.107***
	(0.042)	(0.040)	(0.041)	(0.041)
GAIN$^+$ PEER	−0.044	−0.066	−0.054	−0.061
	(0.046)	(0.046)	(0.047)	(0.048)
LOSS$^+$ PEER	−0.088*	−0.076*	−0.059	−0.041
	(0.045)	(0.046)	(0.047)	(0.047)
GAIN (legacy)	−0.051	−0.052	−0.031	−0.034
	(0.049)	(0.048)	(0.049)	(0.049)
GAIN$^+$ Letter/SMS	−0.081**	−0.081**	−0.083**	−0.072*
	(0.037)	(0.037)	(0.037)	(0.037)
GAIN$^+$ Tag	−0.030	−0.039	−0.051	−0.039
	(0.043)	(0.043)	(0.043)	(0.043)
GAIN$^+$ Letter/SMS Tag	−0.088**	−0.091**	−0.087**	−0.091**
	(0.041)	(0.041)	(0.040)	(0.040)
LOSS$^+$ Tag	−0.070	−0.076*	−0.079*	−0.079*
	(0.043)	(0.043)	(0.042)	(0.043)
LOSS$^+$ PEER Tag	−0.035	−0.041	−0.079*	−0.083*
	(0.047)	(0.048)	(0.047)	(0.048)
Constant	0.234***	0.445***	0.224***	0.217**
	(0.037)	(0.098)	(0.039)	(0.108)
Savings Info	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Fixed Effects	No	No	Yes	Yes
Advisor FE	No	No	No	No
N	1802	1761	1773	1734

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 an interaction between the respective intervention site and the month when the participant is informed about his/her replacement eligibility. Robust standard errors are in parentheses. Significance levels: * : $p < 0.10$, ** : $p < 0.05$, *** : $p < 0.01$.

Appendix C. Calculation of the RRP MVPF

C.1. Concepts

Conceptually, the MVPF for a policy is the ratio of the benefits of the policy to individuals over the net costs of the policy to government.

$$MVPF = \frac{\text{Benefits to individuals}}{\text{Net cost to government}} \tag{2}$$

with appropriate social welfare weights η applied to the benefits for different beneficiaries. The weights η matter substantially for a comparison of the RRP with other public programs. They are essentially irrelevant, however, when ranking design variations that all target the same beneficiaries.

Hahn et al. (2026) show that in the simple case of a policy with subsidy as the only cost, the MVPF of Eq. (2) can be quantified on the basis of four inputs: V, the per-unit externality associated with the target good of the policy x, the price elasticity ϵ of the good x, the price P of the good, and the level of the subsidy τ. This is because Eq. (2) can also be expressed as:

$$MVPF = \frac{1 + \frac{V}{P}(-\epsilon)}{1 + \frac{\tau}{P}(-\epsilon)}. \tag{3}$$

In contrast to this simple case, the RRP is a program that combines two costly measures. One is the encouragement efforts directed at all *eligible* households in the form of the customized information letters and the consultation time during the second visit. Our monetary estimate of these efforts is M per eligible household. Importantly, households do not derive pecuniary benefits from these efforts. Instead, the efforts are intended to lead a high share r of all eligible households to replace their appliances. The second costly measure is the value of the voucher τ paid out to *replacing* households, thereby decreasing the effective price P of the appliance. A marginal change in the RRP budget therefore only delivers a share $\lambda = \tau/(\tau + M/r)$ in the form of a voucher to the eligible household while a share $(1 - \lambda)$ goes to encouragement expenses. Like marketing and nudges in Hahn et al. (2026), budget changes in the RRP therefore have correspondingly smaller impacts on the effective (post-subsidy) price of the target good. The adjusted MVPF is therefore

$$MVPF = \frac{\lambda\left(1 + \frac{V}{P}(-\epsilon)\right)}{1 + \lambda\frac{\tau}{P}(-\epsilon)}, \tag{4}$$

which simplifies to expression (3) for $\lambda = 1$.

C.2. Parametrization

Values for parametrized estimates of expression (4) come from published analyses (Chlond et al., 2025), program data and internal program estimates, and Section 6.

First, the marginal impact of a budget change on the subsidy is estimated at $\lambda = 0.54$, using the voucher value $\tau = 100$ euros, the replacement rate $r = 0.24$ of the baseline (GAIN$^+$) treatment arm, and the RRP's internal estimate of encouragement costs of M =20euros per eligible household.[29] Second, the price elasticity ϵ comes from Chlond et al. (2025) who exploit past subsidy variations in the RRP to determine a mean for ϵ of −1.66 (95-percent confidence interval: −0.37 and −2.94).[30] Third, the ratio of subsidy to purchase price, τ/P comes to 0.18 based on program data that the price per replaced appliance averages $P = 555$ euros.

Fourth, values for V are derived under two mutually exclusive assumptions regarding the so-called "waterbed effect" of the European Emissions Trading System (EU ETS) (Rosendahl, 2019). This effect simply states that in a tradable emissions permit system, policies that do not affect the emissions cap may change the allocation of emissions within the system, but not aggregate emissions of the system. Under one assumption, $V = 0$ because the waterbed effect also applies to the RRP.[31] Under the alternative assumption, favored by the German government, $V > 0$ and V is approximated by the external costs avoided by bringing forward refrigerator replacement through the RRP. The external costs avoided are the social value of non-emitted carbon as a result of electricity savings. To compute the externality reduction per kWh of electricity saved, we require the average CO2 content of German electricity, which is 363gCO2/kWh (Umweltbundesamt, 2024). The German government's

[29] For illustration, assume the program aims to replace 1000 fridges at an aggregate voucher cost of 100.000 euros. Given the replacement rate, this requires preparing, handing out and explaining the information letter to 4348 households at an aggregate marketing cost of 86,957 euros. For the marginal euro going into the program, 0.53 euros therefore go to funding a voucher and 0.47 euros to encouragement efforts.

[30] In Chlond et al. (2025), a voucher variation of 50 euros decreased post-subsidy prices from 372 euros to 322 euros, a decrease of 13.4 percent. The study gives a best estimate of the impact of the variation in post-subsidy prices on demand for refrigerators of 4.9 percentage points, which at its midpoint corresponds to an increase of 22.3 percent in demand for refrigerators.

[31] Emissions reductions in the consumption of electricity on account of the RRP occur in the German power sector. Because this sector is regulated by the EU ETS, reductions there are fully offset by emissions increases elsewhere in what is a supranational multi-sector system. In other words, there is perfect rebound at the system level. This assumption makes the RRP a purely redistributive policy without any environmental externalities whose justification rests on a sufficiently high welfare weight η of the target population only.

preferred estimates of the social cost of carbon (SCC) in 2024 are 300 euros/tCO2eq at 1 percent social discount rate and 880 euros/tCO2eq at zero percent (Umweltbundesamt, 2024). For international comparability and to align with Hahn et al. (2026), we also include the US EPA benchmark estimate of the SCC for 2024 of US$208 or, equivalently, 192 euros (EPA, 2023). These estimates have to be reduced by the price of EU ETS permits, which averaged 80 euros/tCO2 in 2024, to account for the partial internalization. Taken together, these considerations imply from the German government's perspective an externality reduction of 0.080 euros or 0.290 euros per kWh of electricity saved, respectively, and 0.041 euros based on the EPA's SCC.

There are three margins along which an RRP can reduce a household's electricity consumption. The first is bringing forward the age at which the incumbent inefficient refrigerator is replaced. If the RRP induces earlier replacement, then households benefit from lower costs earlier. The second margin is decreasing the age of the replacing refrigerator. The typical replacement among low-income households is a used appliance of 8 to 10 years, as the empirical age distribution encountered in the SSC illustrates (Chlond et al., 2025). Replacement with a new, instead of a used, device delivers higher technological efficiency because the appliance comes from the most recent technological cohort and higher mechanical efficiency because the appliance has not yet undergone wear and tear. The third margin is reducing the energy demand of the replacing refrigerator. The RRP forces the household to buy relatively more efficient refrigerators among the newest cohort of appliances on the market.

Previous analysis (Chlond et al., 2025) does not point to sizable impacts along the first margin: Under the RRP, the age at which refrigerators are commonly replaced is high – 19.1 years on average.[32] This puts the onus of energy savings on the second and third margin. The engineering estimates (Michel et al., 2016; Paul et al., 2022) imply that the second margin, i.e., replacement with a typical newly manufactured appliance instead of an appliance aged 8–10 years, delivers around 35 percent energy savings relative to the existing device.[33] The third margin is the RRP requiring the purchase of a refrigerator of energy efficiency class D or better, rather than the most common classes E and F (Doganoglu et al., 2025). Per EU regulations on energy efficiency classes,[34] this requirement implies an additional 20 percent reduction in energy demand. Taking these effects together means that the new refrigerator requires an estimated 47.5 percent less energy than the used appliance that the household would have ordinarily purchased. Expressed in units of energy, this amounts to a saving of 283.5 kWh per year, based on the energy consumption of used refrigerators in the SSC data. These savings accrue over an expected duration of 9 years.[35] Based on these back-of-the-envelope calculations, undiscounted savings therefore amount to 2552kWh. Taken together, this gives our best estimate of V between 105 euros and 740 euros, depending on the SCC used, and of V/P-ratios between 0.188 and 1.333, respectively.

On the cost side, we have marginal increase in the budget (mechanically set at 1) and the marginal impact on replacement rate, net of the proportional increase in value added tax revenue from voucher spending on new appliances, which is 0.16 for every euro spent.

Table C.1
Marginal value of public funds of the RRP – estimates and sensitivity.

	Waterbed	SCC EPA	SCC, 1%	SCC, 0%
$\epsilon - 2\sigma$	0.524	0.561	0.596	0.784
ϵ	0.475	0.625	0.766	1.528
$\epsilon + 2\sigma$	0.439	0.679	0.907	2.143

Note: Parametrization of expression (4) based on $\lambda = 0.54$ and $\tau/P = 0.18$, $\epsilon - 2\sigma = -.0.37$, $\epsilon = -1.67$, and $\epsilon + 2\sigma = -2.94$. $V/P = 0$ when the waterbed effect is accounted for; otherwise $V/P = 0.19$ for the EPA (2023) Social Cost of Carbon estimate for 2024 under a 2.0 near-term discount rate, $V/P = 0.37$ for the German government's SCC for 2024 under a 1.0 percent discount rate, and $V/P = 1.33$ for its SCC without discounting. Unless otherwise noted, values are drawn from Chlond et al. (2025) and program data.

Table C.1 reports the resulting estimates of the MVPF of the RRP for the baseline treatment (GAIN$^+$).

For the mean elasticity estimate ϵ, the MVPF rises from 0.475 when the waterbed effect is fully accounted for to 1.528 when the waterbed effect is ignored and the German government's upper value of the SCC for 2024 is used. Most of the estimates fall below the unit value. This reflects the high cost of encouragement efforts required to bring households to replace their appliances and collect the subsidy.

While relatively low, the MVPF estimate need not imply that the RRP, an energy efficiency assistance program targeting low-income households, does not represent value for public funds. The redistributive focus would be reflected in the welfare weight η attached to the program (Hahn et al., 2026). Such weights could come from Isaak and Jessen (2024), who connect the MVPF concept to a microsimulation approach to determine the implicit welfare weights enshrined in the German tax code. For household net incomes equivalized for size and composition using the modified OECD scale, these weights can lift the welfare-weighted MVPF above the unit threshold, in particular for the first income decile with a weight of 2.9 (1.4 for the second).

C.3. Ranking of design alternatives

The design variations that the present paper compares differ in their MVPF through differences in the share λ of the marginal change in the budget that goes to replacing households. More precisely, simple algebraic manipulations of expression (4) show that for policies $i, j \in I = \{1, 2, ..N\}$, $MVPF_i > MVPF_j$ whenever $\lambda_i > \lambda_j$, which is the case whenever $r_i/M_i > r_j/M_j$.[36] In other words, policies that achieve a higher replacement rate r at the same "marketing" cost M or the same replacement rate at lower "marketing" cost achieve a higher MVPF.

Fig. A.4 reports replacement rates r_i for each of the eight design variations. 'Marketing' costs M_i per eligible household are the same for all information letter treatments at 20 euros, once the fixed expense of changing the letter has been incurred. For treatments with a reminder component, the RRP estimates costs of 30 euros per eligible household in order to collect and document consent and to implement the reminder in the form of a letter or text message, depending on the household's choice. The higher encouragement expenses of designs with a reminder component therefore need to pay off in the form of higher replacement rates in order to be ranked more highly in the terms of MVPF.

Table 4 reports for each treatment the empirically observed replacement rates, internal estimates of the marketing costs in euros, the

[32] While we lack a clean comparison of the age distribution of refrigerators among the target population in the presence and absence of the RRP, the empirical age at replacement in the presence of the RRP is considerably older than best estimates of the expected age of replacement of 13.0 to 17.5 years for a refrigerator in Germany (Umweltbundesamt, 2025; Laner and Rechberger, 2007). It is also older than the median age of the refrigerators encountered in the total SSC sample (16.0 years, Chlond et al. (2025)).

[33] In terms of technological cohort, past analysis has estimated sales-weighted reductions in energy demand by EU refrigerators of around 25 percent over ten years (Michel et al., 2016). In terms of mechanical age, (Paul et al., 2022) estimate the increase in energy consumption of around 22 percent per ten years.

[34] See Commission Delegated Regulation (EU) 2019/2016.

[35] After 8–10 years, the refrigerators purchased used are then replaced with refrigerators from the originally 'new' cohort, which are by then 8–10 years old.

[36] Note that the parameters V, P, ϵ, and τ are not changed by treatment assignment.

Table D.1
Pooled treatments.

Original treatment	Pooled treatments
(1) control/standard information letter	GAIN (legacy version of the program)
(2) gain frame	GAIN$^+$ (management's new default)
(3) loss frame	LOSS$^+$
(4) personalized gain frame	GAIN$^+$ PEER
(5) personalized loss frame	LOSS$^+$ PEER
(6) gain frame and letter/SMS reminder	GAIN$^+$ REMINDER
(7) gain frame and haptic reminder	GAIN$^+$ REMINDER
(8) loss frame and haptic reminder	LOSS$^+$ REMINDER
(9) personalized loss frame and haptic reminder	LOSS$^+$ PEER REMINDER
(10) gain frame and both SMS/letter and haptic reminder	GAIN$^+$ REMINDER

performance-cost ratio, and the rank of the design alternative among the set of alternatives tested in the RCT. The performance-cost ratio captures the average cost (in euros) per percentage point of replacing households. Across the information letter treatments, the main treatment effect, i.e., changes in the replacement rate, drives the ranking. Treatments involving reminders incur higher marketing costs, most of which go to households that do not replace, and do not raise replacement rates sufficiently, if at all, to make up for the higher costs. As a result, design alternatives that require a reminder are ranked at the bottom.

Appendix D. Evolution of statistical analyses compared to PAP

In this section we describe the main evolution of our statistical analyses compared to the PAP. As already indicated in the AEA RCT preregistration and included pre-analysis plan (https://doi.org/10.1257/rct.10757-2.2), we planned to combine the different reminder treatments into one pooled reminder group for our main analysis in order to avoid issues resulting from low sample sizes. Table D.1 describes how the ten different initial treatment arms are mapped to the eight treatments in our main analysis. In the GAIN domain, we pool the three different reminder types into one GAIN$^+$ REMINDER treatment. In the LOSS domain, by contrast, both the loss frame and the personalized loss frame are combined with the haptic refrigerator tag and kept separate (LOSS$^+$ REMINDER and LOSS$^+$ PEER REMINDER treatments).

As a robustness check, we again differentiate between the different reminder versions (see Tables B.7 and B.8). This procedure does not alter our main observations but helps to explore the mechanism behind lower replacement rates in GAIN$^+$ REMINDER.

We also decided to define the GAIN$^+$ treatment as the omitted treatment category for three different reasons. First, the program management designated this treatment to be the new program baseline. Second, all but our first hypothesis compare treatment variations against this baseline. Consequently, this procedure facilitates the interpretation of the regression results. Third, due to the delay in implementing the GAIN treatment, this group has the fewest observations.

As a last point, in our preferred econometric specification, we do not control for advisor pair fixed effects. The reason is the specific structure of the advisors per visit. While the majority of visits are conducted by advisor pairs (69.33 percent), 30.67 percent of all visits are administered by only one advisor. For such visits, we observe an advisor-1-ID but not an advisor-2-ID. To include advisor-2 fixed effects, we would need to assign a single joint category to such visits with missing advisor-2-IDs to avoid those observations being dropped due to missing values.

As a result, such advisor-2 fixed effects would display a highly unbalanced category structure: 553 of 1803 observations would be assigned to this single imputed category, while the remaining, genuine advisor-2 IDs contain a median of only 7 observations, making this single category roughly 80 times larger than the median genuine category. This is different from the advisor-1 fixed effects, where advisors are not concentrated in just one cell.

However, to still retain the qualitative information of the visit structure, one of our specifications replaces the advisor-2 fixed effect with a dummy distinguishing a genuine advisor-2 ID from the imputed advisor-2 category, while further including the advisor-1 fixed effects.

Appendix E. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jpubeco.2026.105780.

Data availability

The data that has been used is confidential.

References

Abdulai, A., 2023. Information acquisition and the adoption of improved crop varieties. Am. J. Agric. Econ. 105 (4), 1049–1062.

Allcott, H., 2011. Social norms and energy conservation. J. Public Econ. 95 (9–10), 1082–1095.

Allcott, H., 2015. Site selection bias in program evaluation. Q. J. Econ. 130 (3), 1117–1165.

Allcott, H., Greenstone, M., 2024. Measuring the Welfare Effects of Residential Energy Efficiency Programs. NBER Working Paper 23386.

Allcott, H., Kessler, J.B., 2019. The welfare effects of nudges: a case study of energy use social comparisons. Am. Econ. J. Appl. Econ. 11 (1), 236–276.

Andor, M.A., Gerster, A., Peters, J., Schmidt, C.M., 2020. Social norms and energy conservation beyond the US. J. Environ. Econ. Manag. 103, 102351.

Athey, S., 2025. Presidential address: the economist as designer in the innovation process for socially impactful digital products. Am. Econ. Rev. 115 (4), 1059–1099.

Banerjee, A., Hanna, R., Olken, B.A., Satriawan, E., Sumarto, S., Feb 2023. Electronic food vouchers: evidence from an at-scale experiment in Indonesia. Am. Econ. Rev. 113 (2), 514–547.

Barrett, C.B., Harrison, G.W., Jensen, N., Morsink, K., Schneider, M., Swarthout, J.T., Upton, J., 2019. Do no harm: evaluating the welfare effects of behavioral insurance interventions in Ethiopia. J. Econ. Behav. Organ. 162, 167–185.

Barsbai, T., Licuanan, V., Steinmayr, A., Tiongson, E., Yang, D., 2020. Information and the Acquisition of Social Network Connections. NBER Working Paper 27346.

Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L.F., Palmer, C., May 2024. Creating moves to opportunity: experimental evidence on barriers to neighborhood choice. Am. Econ. Rev. 114 (5), 1281–1337.

Bhargava, S., Manoli, D., 2015. Psychological frictions and the incomplete take-up of social benefits: evidence from an IRS field experiment. Am. Econ. Rev. 105 (11), 3489–3529.

Blonz, J.A., 2023. The costs of misaligned incentives: energy inefficiency and the principal-agent problem. Am. Econ. J. Econ. Policy 15 (3), 286–321.

Bonan, J., Cattaneo, C., d'Adda, G., Tavoni, M., 2021. Can social information programs be more effective? The role of environmental identity for energy conservation. J. Environ. Econ. Manag. 108, 102467.

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

Brot-Goldberg, Z., Layton, T., Vabson, B., Wang, A.Y., 2023. The behavioral foundations of default effects: theory and evidence from medicare Part D. Am. Econ. Rev. 113 (10), 2718–2758.

Bruckmeier, K., Riphahn, R.T., Wiemers, J., 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. Empir. Econ. 61, 1567–1616.

Bryan, C.J., Tipton, E., Yeager, D.S., 2021. Behavioural science is unlikely to change the world without a heterogeneity revolution. Nat. Hum. Behav. 5 (8), 980–989.

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

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

Carvalho, L.S., Meier, S., Wang, S.W., 2016. Poverty and economic decision-making: evidence from changes in financial resources at payday. Am. Econ. Rev. 106 (2), 260–284.

Chapman, J., Snowberg, E., Wang, S.W., Camerer, C., 2025. Looming large or seeming small? Attitudes towards losses in a representative sample. Rev. Econ. Stud. 92 (5), 2893–2922.

Chen, C., 2023. Framing energy-efficiency programs: a survey experiment. Energy Policy 183, 113776.

Chlond, B., Goeschl, T., Kesternich, M., 2025. Money versus procedures — evidence from an energy efficiency assistance program. J. Environ. Econ. Manag. 130, 103080.

Composto, J.W., Weber, E.U., 2022. Effectiveness of behavioural interventions to reduce household energy demand: a scoping review. Environ. Res. Lett. 17 (6), 063005.

Davis, L.W., Fuchs, A., Gertler, P., 2014. Cash for coolers: evaluating a large-scale appliance replacement program in Mexico. Am. Econ. J. Econ. Policy 6(4), 207–238.

DeGolia, A.H., Hiroyasu, E.H., Anderson, S.E., 2019. Economic losses or environmental gains? Framing effects on public support for environmental management. PLoS One 14 (7), e0220320.
Degtiar, I., Rose, S., 2023. A review of generalizability and transportability. Annu. Rev. Stat. Appl. 10, 501–524.
Doganoglu, T., Grzybowski, L., Verboven, F., 2025. The Impact of the New EU Energy Label 2021 on Energy Consumption of Domestic Appliances. Working Paper 26/2025. Faculty of Economic Sciences, University of Warsaw.
Domurat, R., Menashe, I., Yin, W., 2021. The role of behavioral frictions in health insurance marketplace enrollment and risk: evidence from a field experiment. Am. Econ. Rev. 111 (5), 1549–1574.
Duflo, E., Glennerster, R., Kremer, M., 2007. Using randomization in development economics research: a toolkit. In: Handbook of development economics, vol. 4. pp. 3895–3962.
EPA, 2023. EPA Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances.
Ericson, K.M., 2017. On the interaction of memory and procrastination: implications for reminders, deadlines, and empirical estimation. J. Eur. Econ. Assoc. 15 (3), 692–719.
Escobar, L., Pedraza, A., 2023. Active trading and (poor) performance: the social transmission channel. J. Financ. Econ. 150 (1), 139–165.
Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., Staake, T., 2023. Complementarities in behavioral interventions: evidence from a field experiment on resource conservation. J. Public Econ. 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 (Accessed: 21 November 2023).
Federal Statistical Office of Germany, 2023b. Tabelle 61243-0001: Strompreise für Haushalte: Deutschland, Halbjahre, Jahresverbrauchsklassen, Preisarten.
Federal Statistical Office of Germany, 2023c. Umweltökonomische Gesamtrechnungen. Stromverbrauch der privaten Haushalte nach Haushaltsgrößenklassen. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/private-haushalte/Tabellen/stromverbrauch-haushalte.html (Accessed: 30 October 2023).
Federal Statistical Office of Germany, 2023d. Wohnfläche. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Wohnen/Tabellen/tabelle-wo4-wohnflaeche.html (Accessed: 21 November 2023).
Fehr, D., Fink, G., Jack, B.K., 2022. Poor and rational: decision-making under scarcity. J. Polit. Econ. 130 (11), 2862–2897.
Finkelstein, A., Notowidigdo, M.J., 2019. Take-up and targeting: experimental evidence from SNAP. Q. J. Econ. 134 (3), 1505–1556.
Fowlie, M., Greenstone, M., Wolfram, C., 2015. Are the non-monetary costs of energy efficiency investments large? Understanding low take-up of a free energy efficiency program. Am. Econ. Rev. 105 (5), 201–204.
Fowlie, M., Greenstone, M., Wolfram, C., 2018. Do energy efficiency investments deliver? Evidence from the weatherization assistance program. Q. J. Econ. 133(3), 1597–1644.
Goeschl, T., 2019. Cold case: the forensic economics of energy efficiency labels for domestic refrigeration appliances. Energy Econ. 84, 104468.
Gonzales, M.H., Aronson, E., Costanzo, M.A., 1988. Using social cognition and persuasion to promote energy conservation: a quasi-experiment. J. Appl. Soc. Psychol. 18 (12), 1049–1066.
Gopalan, M., Pirog, M.A., 2017. Applying behavioral insights in policy analysis: recent trends in the United States. Policy Stud. J. 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., Langetieg, P., Manoli, D., Payne, M., Schafer, B., Sebastiani, M., 2017. Reminders and recidivism: using administrative data to characterize nonfilers and conduct EITC outreach. Am. Econ. Rev. 107 (5), 471–475.
Hahn, R.W., Metcalfe, R.D., 2021. Efficiency and equity impacts of energy subsidies. Am. Econ. Rev. 111 (5), 1658–1688.
Hahn, R.W., Hendren, N., Metcalfe, R.D., Sprung-Keyser, B., 2026. A welfare analysis of policies impacting climate change. Am. Econ. Rev. 116 (7), 2368–2421.
Hallsworth, M., 2023. A manifesto for applying behavioural science. Nat. Hum. Behav. 7 (3), 310–322.
Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I., 2017. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. J. Public Econ. 148, 14–31.
Halpern, D., Sanders, M., 2016. Nudging by government: progress, impact & lessons learned. Behav. Sci. Policy 2 (2), 53–65.
Haushofer, J., Fehr, E., 2014. On the psychology of poverty. Science 344 (6186), 862–867.
Hendren, N., Sprung-Keyser, B., 2020. A unified welfare analysis of government policies. Q. J. Econ. 135 (3), 1209–1318.
Holzmeister, F., Huber, J., Kirchler, M., Schwaiger, R., 2022. Nudging debtors to pay their debt: two randomized controlled trials. J. Econ. Behav. Organ. 198, 535–551.
Homar, A.R., Cvelbar, L.K., 2021. The effects of framing on environmental decisions: a systematic literature review. Ecol. Econ. 183, 106950.
Hotard, M., Lawrence, D., Laitin, D.D., Hainmueller, J., 2019. A low-cost information nudge increases citizenship application rates among low-income immigrants. Nat. Hum. Behav. 3 (7), 678–683.
Houde, S., Aldy, J.E., 2017. Consumers' response to state energy efficient appliance rebate programs. Am. Econ. J. Econ. Policy 9(4), 227–255.
Isaak, N., Jessen, R., 2024. Normative Judgments Implicit in the Tax System: A Simulation Approach. Technical report, IZA Discussion Papers No. 17566.
Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2016. Getting to the top of mind: how reminders increase saving. Manag. Sci. 62 (12), 3393–3411.
Khanna, T.M., Baiocchi, G., Callaghan, M., Creutzig, F., Guias, H., Haddaway, N.R., Hirth, L., Javaid, A., Koch, N., Laukemper, S., et al., 2021. A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings. Nat. Energy 6 (9), 925–932.
Kószegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.
Kühberger, A., 1998. The influence of framing on risky decisions: a meta-analysis. Organ. Behav. Hum. Decis. Process. 75 (1), 23–55.
Laibson, D., List, J.A., 2015. Principles of (behavioral) economics. Am. Econ. Rev. 105 (5), 385–390.
Laner, D., Rechberger, H., 2007. Treatment of cooling appliances: interrelations between environmental protection, resource conservation, and recovery rates. Resour. Conserv. Recycl. 52 (1), 136–155.
Linos, E., Prohofsky, A., Ramesh, A., Rothstein, J., Unrath, M., 2022. Can nudges increase take-up of the EITC? Evidence from multiple field experiments. Am. Econ. J. Econ. Policy 14 (4), 432–452.
List, J.A., 2020. Non Est Disputandum de Generalizability? a Glimpse Into the External Validity Trial. NBER Working Paper 27535.
List, J.A., 2024. Optimally generate policy-based evidence before scaling. Nature 626 (7999), 491–499.
List, J.A., Shaikh, A.M., Xu, Y., 2019. Multiple hypothesis testing in experimental economics. Exp. Econ. 22 (4), 773–793.
Löschel, A., Rodemeier, M., Werthschulte, M., 2023. Can self-set goals encourage resource conservation? Field experimental evidence from a smartphone app. Eur. Econ. Rev. 160, 104612.
Mani, A., Mullainathan, S., Shafir, E., Zhao, J., 2020. Scarcity and cognitive function around payday: a conceptual and empirical analysis. J. Assoc. Consum. Res. 5 (4), 365–376.
Michel, A., Attali, S., Bush, E., 2016. Energy Efficiency of White Goods in Europe: Monitoring the Market with Sales Data – Final Report. ADEME.
Mullainathan, S., Shafir, E., 2013. Scarcity: Why Having Too Little Means so Much. Macmillan.
OECD, 2025. Mind Shift, Green Lift: Six Behavioural Science Trends for Environmental Policy. OECD Public Governance Policy Papers. Organization of Economic Cooperation and Development, Paris..
Park, J., Son, W., Moon, H., Woo, J., 2023. Nudging energy efficiency behavior: the effect of message framing on implicit discount rate. Energy Econ. 117, 106485.
Paul, A., Baumhögger, E., Elsner, A., Reineke, M., Hueppe, C., Stamminger, R., Hoelscher, H., Wagner, H., Gries, U., Becker, W., et al., 2022. Impact of aging on the energy efficiency of household refrigerating appliances. Appl. Therm. Eng. 205, 117992.
Pearl, J., Bareinboim, E., 2011. Transportability of causal and statistical relations: a formal approach. In: Proceedings of the AAAI Conference on Artificial Intelligence, 25, pp. 247–254.
Rosendahl, K.E., 2019. EU ETS and the waterbed effect. Nat. Clim. Change 9 (10), 734–735.
Silvi, M., Padilla Rosa, E., 2021. Reversing impatience: framing mechanisms to increase the purchase of energy-saving appliances. Energy Econ. 103, 105563.
Stojanovski, O., Leslie, G.W., Wolak, F.A., Wong, J.E.H., Thurber, M.C., 2020. Increasing the energy cognizance of electricity consumers in Mexico: results from a field experiment. J. Environ. Econ. Manag. 102, 102323.
Sussman, R., Chikumbo, M., Gifford, R., 2018. Message framing for home energy efficiency upgrades. Energy Build. 174, 428–438.
Umweltbundesamt, 2024. Methodological Convention 3.2 for the Assessment of Environmental Costs: Value Factors.
Umweltbundesamt, 2025. Ökologische und ökonomische Vergleichsrechnung von Haushaltsgeräten. Recherchegrundlagen, Modellierung und Auswertung.