Energy Policy 169 (2022) 113156

Contents lists available at ScienceDirect

Energy Policy

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

Heat pump inspections result in large energy savings when a pre-selection of households is performed: A promising use case of smart meter data

Andreas Weigert [a], Konstantin Hopf [a], Sebastian A. Günther [a], Thorsten Staake [a,b,*]

[a] Department of Information Systems and Applied Computer Sciences, University of Bamberg, An der Weberei 5, 96047 Bamberg, Germany

[b] Department of Management, Technology and Economics, ETH Zurich, Weinbergstr. 56/58, 8092 Zurich, Switzerland

ARTICLE INFO

Keywords:
Heat pump
Energy consulting
Energy efficiency
Smart meter data
Data analytics

ABSTRACT

Heat pumps play an important role in the energy transition. They can extract renewable energy from the air or ground and increasingly replace fossil heating systems in buildings. In operation, however, heat pumps often consume more electricity than necessary due to incorrect settings and installation deficiencies. Although many setting and installation issues are easy to fix, problems often go unnoticed, and the saving potential from quick fixes remains unclear. In a study with 297 Swiss households (41 treatment, 256 control) running for four years, we investigated an energy efficiency campaign in which the treatment group received a professional heat pump inspection and user training. We found considerable heterogeneity with respect to the savings achieved. We derived two criteria based on smart meter data that enable utilities to identify relevant households and thus boost the impact of such efficiency campaigns: For example, pre-selecting half of the households based on available information results in average savings of 1,805 kWh (15.2%) per year and household in the high-potential group compared to no savings in the low-potential group. Thus, heat pump inspections among pre-selected households can lead to large, cost-effective electricity savings, and we show that common smart meter data makes such pre-selection feasible.

1. Introduction

Fossil-fuelled heating systems satisfy one of the most basic needs of our society while at the same time contributing significantly to the emission of greenhouse gases. Almost a quarter of the world's final energy consumption is attributable to space heating, with fossil fuels still accounting for the lion's share of the energy used (IEA, 2019). The use of fuels causes direct emissions from burning (e.g. CO_2, NO_x and particulate matter) and indirect emissions (e.g. CH_4) that arise even for biofuels during their production, transportation, and storage (Adams et al., 2015; Lauvaux et al., 2021).

Efforts to reduce these emissions have been intensified for many years by replacing fossil heating systems with such that rely on renewable sources. Among those, heat pumps have received considerable attention, as they can use thermal energy from the ground or ambient air. From a technical perspective, a heat pump can extract with one unit of electricity invested three to five times more useable energy (IEA, 2019). Modern systems even operate without problematic greenhouse gases as refrigerants and have an excellent energy balance over the entire life cycle. Heat pumps also gained economic attractiveness as investment costs were reduced by economies of scale and by policy-makers granting subsidies (IEA, 2021). On the other hand, the relative operating costs—compared to traditional heating systems—have fallen in countries that have introduced carbon prices for fossil fuels and due to increases in the gas market (The Economist, 2021).

For these reasons, heat pumps are often the first choice for new buildings. In 2018 almost 50% of the newly-built U.S. multi-family houses and over 40% of single-family homes were equipped with heat pumps (IEA, 2020), compared to 43% in Germany (Breisig et al., 2020) and 90% for 2019 in Switzerland (Renz, 2020). In some Scandinavian countries, heat pumps already account for 30% of the residential building stock, and European sales indicate a robust overall market expansion (EHPA, 2019; IEA, 2019).

However, this desirable development creates additional electricity demand and peaks. Both must be satisfied by producers of (renewable) electricity and managed by grid operators. In an ongoing debate, several researchers evaluate the extent to which grids can cope with a high heat pump penetration. For example, the peak capability of the UK's grid

* Corresponding author.
E-mail addresses: andreas.weigert@uni-bamberg.de (A. Weigert), konstantin.hopf@uni-bamberg.de (K. Hopf), sebastian.guenther@uni-bamberg.de (S.A. Günther), thorsten.staake@uni-bamberg.de (T. Staake).

https://doi.org/10.1016/j.enpol.2022.113156
Received 6 October 2021; Received in revised form 16 June 2022; Accepted 4 July 2022
Available online 20 August 2022
0301-4215/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

must be increased by 14% (60%) for penetration rates of 20% (50%) (Eggimann et al., 2019; Love et al., 2017). Taking a more holistic view that considers the grid of European countries for a full-rollout scenario, Connolly (2017) estimates that peak demand could triple, while others expect that electricity infrastructure in many but not all countries can handle the additional load (Thomaβen et al., 2021). Even modern heat pumps can increase the stress on the grid since they can achieve high output temperatures comparable to conventional systems and are now more likely to be installed in the more energy-hungry building stock (EHPA, 2019). Their deployment must therefore be accompanied by widespread investments in the efficiency of buildings (Jack et al., 2021) but is then often offset by growing residential space (Ürge-Vorsatz et al., 2015).

Alongside important factors like fabric-first retrofits to increase the efficiency of buildings, grid capability, the supply of green electricity, and a fast spread of heat pumps, their efficiency in the field becomes one of the key factors to transforming space heating successfully. However, one practical challenge is that heat pumps often fail to meet minimum efficiency standards or operate well below the manufacturer's stated efficiency, resulting in higher-than-necessary electricity consumption (Caird et al., 2012; Qiao et al., 2020; Yin et al., 2019). This gap between nominal and real efficiency is well known. Many causes of poor performance have been documented in the literature, including planning and design flaws, installation faults, quality-related problems of components, non-optimal or even incorrect configuration of parameters, and users with little knowledge of the technology (Domanski et al., 2014; Madani and Roccatello, 2014; Miara et al., 2011, 2017; Qiao et al., 2020; Winkler et al., 2020). An extensive stream of research proposes approaches for the next generation of heat pumps to improve their mechanical design, dynamic controls, simplicity of installation, and usability. Many proposals rely on AI-based control algorithms, networking capabilities, and more sensor information, thus targeting improvements to the next generation of heat pumps.

In the research presented here, we set a different focus: We target the large number of installed systems, which will often be in use for another decade, and suggest how to improve heat pump inspection campaigns that focus on easy-to-fix problems in the existing stock. The measures are also relevant to heat pumps currently on the market, which often do not yet have the latest control technology nor use Internet connectivity.

There are several easily correctable causes of low performance, such as less than optimal parameter configuration and lack of user knowledge (Caird et al., 2012; Miara et al., 2011, 2017; Qiao et al., 2020). One example is an overcautious heating curve setting that may occur due to split incentives in selecting heating systems settings. Installers tend to set the flow temperature rather too high to ensure comfort for the residents (and fewer requests for unpaid corrections) at the expense of higher energy consumption on the resident's side. In such cases, residents tend to adjust the room temperature and do not modify the heating control settings, which leads to overall lower system efficiency. Such maladjustments often go unnoticed, even if heating professionals can immediately remedy such easy-to-fix problems in a short inspection visit. Although it appears that easy-to-solve performance issues offer great potential, it remains unclear how much energy households save on average if they receive professional on-site heat pump inspections along with user training. Thus, the first research question (RQ) of our study is:

> **RQ1:** How much do on-site energy efficiency inspections for heat pumps decrease a household's electricity consumption?

Answering RQ1 allows for assessing the cost-effectiveness and carbon abatement cost of the corresponding heat pump services.

Previous energy efficiency campaigns—and our own analysis—indicate that the saving effects of on-site inspections vary widely. For example, Taylor et al. (2014) report that only households with a comparably low a priori energy performance showed significant savings after receiving an energy audit, including a full walk-through inspection. Even between households with similar problems (e.g. unnecessary heating in summer, additional electric heating), the savings potential can differ significantly (Miara et al., 2011). Moreover, Domanski et al. (2014) found diverse interactions of problems: While there are problem constellations in which efficiency losses accumulate as expected, there are also constellations in which minor changes occur compared to the single fault condition and others in which the performance even deteriorates. Therefore, we investigate the variance of energy savings among the households in our study that received a professional heat pump inspection and user training in an attempt to answer the following RQ:

> **RQ2:** How much do the savings differ within different subgroups of households?

The answer to RQ2 provides insights into whether and to what extent a focus on specific groups of households can increase the performance of such efficiency programs.

Beyond descriptive analyses, we also obtain decision-relevant insights for utility companies. Although previous studies (Caird et al., 2012; Gleeson and Lowe, 2013; Miara et al., 2011, 2017; Qiao et al., 2020; Yin et al., 2019) named important stakeholders like installers, planners, manufacturers, and users together with recommendations for actions to contribute to a higher performance of heat pumps, we point to the critical role of utilities, as they already started to execute several energy efficiency mandates (e.g. EU, 2018, 2012) and have experience in implementing energy efficiency campaigns. They are also investing in expensive smart metering infrastructures, which generates data streams on the electricity consumption of households, including the consumption of heat pump installations (EC, 2014). Smart metering can enable heat pumps to increase grid stability and facilitate the integration of renewable energy sources (Fischer and Madani, 2017; Hillberg et al., 2019), for example, using dynamic electricity tariffs (Gottwalt et al., 2011). However, utilities are still looking for new use cases for leveraging their smart meter infrastructure to get a return on their significant investments. (Ecoplan, 2015).

Efficiency campaigns—whether mandated by regulators or launched by companies because they help them to attract new customers—can win significantly if high-potential participants are deliberately selected in advance (Taylor et al., 2014). Thus, our third RQ investigates selection criteria to identify households with high saving potential:

> **RQ3:** Can easy-to-obtain information help to identify high-potential households?

The answer to RQ3 will inform the reader how to implement segmentation to identify targets with a high achievable gain and provide concrete data from a large case on how this boosts savings and cost-effectiveness.

To answer the three RQs, we examined an energy efficiency campaign from a Swiss utility that offers professional on-site inspections of residential heat pumps, including remediation of easy-to-solve problems and user training. In a quasi-experiment, we examined monthly electricity consumption data from smart meters that measured heat pump consumption together with weather data from 297 households. While 41 households received a heat pump inspection, 256 households with a heat pump from another utility (that did not offer such a service) served as a control group. In total, we analysed data over an average period for each household of four years, yielding 14,815 data points.

The remainder of this article is organised as follows: Section 2 reviews previous studies on the performance of heat pumps in the field, causes for low performance and approaches to detect faulty heat pumps. Section 3 describes the field experiment and data used in this analysis. Section 4 describes the analysis methods used, presents and discusses the results. Finally, in Section 5, we summarise our findings, conclude implications for policy and utilities, and name the limitations of this study together with suggestions for future work.

2. Background and related work

The performance of heat pumps attracted significant attention from researchers. Several studies focus on the engineering of energy-efficient heat pumps (e.g. Deng et al., 2020) and their combination with other heating systems (e.g. Long et al., 2021; Pater, 2019). The primary rationale of these studies is to enhance future heat pump technology. The focus of our research, by contrast, lies in the performance improvements during the long operation phase of those installations. It is the most energy-intense part of the entire heat pump life cycle and accounts for 50% of their total environmental impact (Marinelli et al., 2019). Keeping in mind this differentiation, we review the literature on the performance measurement of heat pumps, causes of low heat pump performance and their easiness to fix during the operation phase, the impact of easy-to-solve interventions, and ways to recognise faulty heat pump installations.

2.1. Heat pump performance

Policymakers[1] introduced standardised methods to measure the performance of heat pumps in the laboratory (e.g. test standards of heat pumps EN14511, calculation standard SCOP EN14825) and defined minimum system efficiencies (EC, 2013a, 2009). These standards are primarily designed to improve the comparability of space and water heating systems on the market. They suggest energy efficiency labelling (Regulation EU 1369/2017 and CDR EU 811/2013) to foster the implementation of more efficient systems in their class (EC, 2013b; EP, 2017). However, they are not guiding to estimate the actual performance in the field, which we describe below.

In the laboratory, the performance of heat pumps is often measured using the coefficient of performance (COP), which is defined as the ratio of heat energy produced (Q) to the electrical energy consumed (W) from the heat pump $COP = Q / W$. It is important to note that W typically only includes the energy used by the heat pump, together with fans or necessary auxiliary pumps. It does not include additional consumers that belong to the heating system (e.g. circulator pumps, buffer storage, electrical backup heater) and the hot water production system (e.g. hot water storage, hot water circulation pump) (Gleeson and Lowe, 2013). During the COP measurement, the input temperature of the used heat reservoir is constant. For example, in the case of air source heat pumps, one assumes an input air temperature of 2 °C or 7 °C with an output flow temperature of 35 °C. This performance evaluation is typically applied once for each model in a test laboratory but is not guiding in estimating the actual performance in the field. The heat pump in operation has varying input and output temperatures. In addition, it is always part of a more complex configuration of several systems that belong to the heat pump, such as the individual heat distribution and hot water production systems, together with appropriate settings (e.g. operation cycles, temperature levels).

Unsurprisingly, several studies showed that the actual performance of residential heat pumps in the field is often below the expected metrics. Table 1 gives an overview of studies focusing on residential installations in different markets and periods that reported the underperformance of air source heat pumps (ASHP) and ground source heat pumps (GSHP). For example, Qiao et al. (2020) found that 60% of the investigated systems in China operate only in the least efficient class accepted by the regulations, while 10% even failed to fulfil the minimum energy efficiency standard. Likewise, Caird et al. (2012) found that in the UK, "only a few heat pumps reached the minimum system efficiency to count as renewable energy under the EU renewable directive" (p. 285). A similar picture emerges in the U.S., where 80% of the investigated ground source heat pumps studied operate below their nominal efficiency level (Yin et al., 2019). In a smaller study for ground source heat pumps, Puttagunta et al. (2010) note that "the overall winter system COP […] was 3.6; the advertised unit COP is 5.0." (p. 949). Finally, in a recent study, Chesser et al. (2021) found that "The majority of ASHPs underperformed. The level of difference between the average predicted COP and the manufacturers for the 8.5 kW ASHP is −16% […] at an outtemp of 7 °C […] and for the 11.2 kW ASHP –24%" (p. 9–10).

Table 1
Overview of studies reporting underperformance of heat pumps in the field.

	Puttagunta et al. (2010)	Caird et al. (2012)	Yin et al. (2019)	Qiao et al. (2020)	Chesser et al. (2021)
Sample size	3	83	32	28	12
Buildings					
Residential	x	x	x	x	x
Commercial	–	–	–	x	–
Privately owned	x	x	x	–	x
Social housing	–	x	–	–	–
Location	USA	UK	USA	China	Ireland
Heat pump					
ASHP	–	x	–	–	x
GSHP	x	x	x	x	–
Performance[A]	GSHP 3.3–4.3	ASHP 1.2–3.2	Horizontal GSHP 1.2–4.6	Soil source heat pumps 4.3	ASHP 3.0–3.6 (COP at 7 °C)
				GSHP 4.2	
		GSHP 1.3–3.3	Vertical GSHP 2.0–3.8	Surface water heat pumps 4.2	

Notes: [A] The authors use different definitions of the COP. Some include only the electricity consumption of the heat pump (e.g. Qiao et al., 2020), and some also use the consumption of auxiliary devices, which limits the comparability. The difficulty of comparing performance metrics from different studies was also reported by Gleeson and Lowe (2013).

2.2. Causes for low heat pump performance and solution strategies

Motivated by the fact that heat pump performance worldwide often lies below the minimum efficiency standards or the manufacturer's stated efficiency (Caird et al., 2012; Chesser et al., 2021; Gleeson and Lowe, 2013; Puttagunta et al., 2010; Qiao et al., 2020; Yin et al., 2019), several studies investigate the causes of low heat pump performance and give recommendations to remedy them. The literature names (1) planning and design flaws, (2) quality problems with heat pump components, (3) installation faults, (4) configuration of parameters, as well as (5) poorly trained users of heat pumps. We describe these categories below, name examples of issues, and discuss how complex they are to fix once the system is installed.

First, planning and design flaws include, for example, wrong system capacity and wrong sizing of the ductwork (Decuypere et al., 2022; Domanski et al., 2014; Miara et al., 2011, 2017). Heat pumps are more likely oversized if poorly insulated buildings with higher thermal demand are retrofitted and when the budget is limited (Decuypere et al., 2022). An explanation for oversizing might be that homeowners tend to order energy-related investments with urgency. For example, replacing the broken heating system might be favoured over "less necessary" prerequisite investments in the building envelope (e.g. new windows, insulation of walls) and thus compensated with a larger heat pump. Although professionals can often easily identify such problems afterwards, their remediation requires significant labour and is usually not economical. Literature suggests for future installations that only well-trained and certified planners and installers should carry out heat

[1] We review EU's policies here as examples, similar policies are introduced in other markets.

pump projects to avoid such planning and design flaws (Caird et al., 2012; Gleeson and Lowe, 2013). In addition, effective coordination among contractors can also prevent such problems (Caird et al., 2012; Decuypere et al., 2022).

The second cause of low heat pump performance is quality problems of components (Madani and Roccatello, 2014; Miara et al., 2011, 2017; Qiao et al., 2020). Two examples were frequently named: Valves that do not close properly and lead to a discharge in the hot water storage and faulty temperature sensors, which can cause severe problems in the defrosting cycle of heat pumps (Madani and Roccatello, 2014; Miara et al., 2011, 2017). Manufacturers can avoid this class of issues by improving the quality management for component production processes, which comes at a low cost compared to the overall price (Madani and Roccatello, 2014). The exchange of low-quality components in an operating system can also be associated with comparably high effort.

Third, installation faults can occur even when a heat pump is properly planned and its components work properly. Examples are refrigerant overcharge or undercharge, improperly mounted sensors, and poorly insulated pipes (Domanski et al., 2014; Miara et al., 2011, 2017; Winkler et al., 2020). Experts cannot solve many of these issues immediately but must first order parts or refrigerants.

The fourth set of problems is non-optimal parameter configurations. Issues occur in connection with the heating curve (e.g. flow temperature too high), the operation cycle (e.g. night setback), the charge pump (e.g. speed regulation), electrical backup heaters (e.g. unintended use), and the domestic hot water tank (e.g. too high temperature) (Miara et al., 2011, 2017). Another reason might be the complexity of a system. Using legacy heating distribution systems (i.e. radiators) not only decrease the overall performance since they were designed for high-temperature heating systems, but they are sometimes used altogether with underfloor heating (Caird et al., 2012). This leads to several heating circuits that are more complex to be optimised. The same logic applies to an external domestic hot water system. This category is therefore well suited for low-cost and quick improvements.

Finally, several papers cite heat pump users' lack of knowledge in addition to the technical problems. It seems that there is a strong dependency between systems with higher efficiency and users with general knowledge and understanding of their heat pump system (Caird et al., 2012; Miara et al., 2011, 2017; Qiao et al., 2020). Well-trained users, for example, do not turn their heating system down for shorter absences (only for longer vacations) and let the heat pump switched on overnight, which is contrary to how to operate traditional heating systems (Caird et al., 2012). Knowledge of the heating system and the electricity tariffs can also avoid unnecessary costs by shifting the operation cycle of the heating system into cost-optimal times (e.g. night and day rates). Training is also relevant from a social perspective since one study indicates that social housing residents had less knowledge of heat pumps than private households (Caird et al., 2012). Likewise, it was shown that tenants with low levels of education who use heat pumps could suffer from energy poverty and thus heat their homes only as much as they can afford (Primc et al., 2019). Enhancing the training of heat pump users can thus decrease heat pump consumption and reduce costs without reducing thermal comfort.

The last two categories can be remedied with comparably easy-to-solve measures to improve the heat pump performance. Experts can carry them out within a single on-site visit and without any component replacement. Both categories are also related since users' awareness can lead to an optimised configuration (e.g. activate backup heater only if necessary). Measures that can be directly implemented and enhance the users' knowledge are also likely to lead to immediate performance gains. In opposite, measures that fall into the category "recommended for future implementation" since they require significant follow-up effort (i.e. modification of the installation or change of components) are often not timely implemented by owners (Barbetta et al., 2015; Broberg et al., 2019; Murphy, 2014). It is therefore helpful to focus on quick-win categories.

2.3. *Impact of easy-to-solve interventions on heat pump performance*

Despite the potential of such easy-to-solve measures, only a few studies explicitly investigate the impact of such interventions. Miara et al. (2011, 2017) found that solving severe configuration issues can increase heat pump performance. Other authors document a correlation between user knowledge and efficiency (Caird et al., 2012; Qiao et al., 2020). So far, it is still unclear how much energy conservation is possible through user training and fixing easy-to-solve configuration issues.

2.4. *Identification of heat pumps with saving potential*

Despite the so far cited problems with heat pump performance in the field, many heat pumps are actually operating properly. When focusing on planning impactful energy efficiency interventions, campaigns should focus on those installations that have a high need to be inspected since this can significantly increase the effectiveness of such campaigns (Taylor et al., 2014). Existing approaches to identifying heat pump installations with a high saving potential offer little guidance, as described below.

Studies investigating the causes of low heat pump performance, which we reviewed in the previous section, do not offer distinct selection approaches to identify installations with a low performance from operational data. Yet, an increasing field of research uses statistical or machine learning techniques to identify characteristics of heat pumps based on sensor data after they have been installed. Some studies identify faulty conditions, such as poor refrigerant charge levels (Bode et al., 2020; Eom et al., 2019) and use data from multiple sensors (e.g. temperatures, pressures, and electricity consumption) placed in and around the heat pump. For practical application, the availability of such sensor data might be the biggest obstacle because sensors are mainly proprietary and existing heat pumps are usually not equipped with sufficient communication interfaces for data exchange. Furthermore, Bode et al. (2020) reported transferability problems when they applied models based on laboratory data in the field.

Other works focus on approaches that make use of smart metering technology, which can measure the electricity consumption of heat pumps alone or together with the entire household consumption, at different time intervals, for example, at daily or 15-minute measurements, and to communicate the consumption traces to utilities (EC, 2014). In contrast to the previously mentioned proprietary heat pump sensors, smart metering is already installed in many European buildings today, and a roll-out of 77% is expected by 2024 (Tounquet and Alaton, 2020). Several studies demonstrate how smart meter data can be used with weather data to predict the heat pump existence in buildings (Fei et al., 2013; Hopf et al., 2018) and even more detailed characteristics such as the heat pump type and age (Weigert et al., 2020). However, to the best of our knowledge, there are no studies examining smart meter data available to utility companies that allow identifying underperforming heat pumps.

We thus see utilities as an important stakeholder group to apply a data-driven selection of underperforming heat pumps in efficiency campaigns and thus contribute to a higher heat pump efficiency in the field. Furthermore, utilities often have a public mandate to reduce their customers' electricity consumption. Efficiency measures, like the heat pump inspections we analyse in this study, are usually not fully paid by customers who order the service but are co-financed by the public. Utilities redistribute ratepayers' charges as subsidies, that are, for example, raised by additional taxes on the electricity price under public mandate (Alberini and Towe, 2015; Cho et al., 2019; EU, 2021; Taylor et al., 2014) or also use their financial reserves (energieschweiz, 2015). In such constellations, utilities must ensure to implement related measures cost-effectively and that profiteers of such measures make their fair contribution. This makes it necessary to determine the effectiveness of such measures in terms of saved energy, cost, and benefit.

One challenge with many data-based approaches that aim to identify promising candidates for energy efficiency campaigns is the regression-to-the-mean effect. This statistical effect means that an extreme measurement (under random influences and with a high correlation in the data) is relativised by measurements in consecutive periods (Barnett, 2004). For the case of selecting households for a treatment, this means that households with a rather high or low energy consumption in an observation period tend to have consumption in future periods that is, on average, closer to a "normal" consumption (e.g. a household heated more in one period because of staying longer at home than usually). This effect was observed in several energy conservation campaigns (e.g. Verkooijen et al., 2015; Viggers et al., 2019) and could lead to overestimated treatment effects. Consequently, if selection criteria based on consumption characteristics are used to identify promising candidates for efficiency campaigns, the research design must control for this effect.

3. Research approach

We obtained data from two electricity utilities in Switzerland. The data from utility A includes electricity meter readings and inspection reports from 41 households that received a professional on-site heat pump inspection, which included fixes of easy-to-solve problems and user training. Utility B's data set provides electricity consumption data from 256 households with comparable heat pump installations that did not receive a professional on-site heat pump inspection. We first describe the conducted intervention and the study design of our quasi-experiment. Then, we introduce the electricity consumption and weather data used in our analysis.

3.1. Experimental setting

Utility A conducted the studied heat pump inspections as a customer service. This service was priced at CHF 400 (ca. USD 426), while customers in their core sales area received a 50% discount as they already paid indirectly for the service via a surcharge on the electricity price for such services. The inspection aimed to improve the performance of the households' heating system and included an on-site visit of a professional energy efficiency consultant who evaluated the overall performance of the entire heating system, for example, the heat pump, thermal storage, pipes, and radiators. When necessary, the expert solved easy-to-fix problems, optimised parameters on the heating control unit, and trained users on how to operate the system properly in case users had issues with the system (e.g. too low thermal output from the heat pump). The consultant documented the evaluation together with recommendations for repairs that cannot be remedied immediately and should be carried out by an installer afterwards in a questionnaire-like report. The consumer received the report at the end of the inspection.

From over 200 households that have booked the inspection service via the utility's website between 2015 and 2020, the electricity consumption data from 41 households were suitable for our analysis (as we describe in the next section). We refer to these households hereafter as the *treatment group*. Given that utility A could not provide similar electricity consumption data on households with heat pumps that have not received a heat pump inspection,[2] we used a data set on households with heat pumps from utility B, which serves a very close geographic sales area (the average distance between the households in both data sets is 51.6 km). This data set stemmed from a joint research project with utility B, which offered no such service at that time. We used these households as a *control group* in our study.

Our study investigated the heat pump inspection interventions based on these two data sets in a quasi-experimental design. The intervention took place for each household of the treatment group at an individual point in time. Therefore, we divided the data into the period before the intervention, which we refer to as the *baseline phase*, and the time after as the *treatment phase*. The start of the baseline phase was individual for each household, as the smart meter roll-out was ongoing at both utilities, and households were gradually equipped with smart meters during our study period. Due to the different start times and time periods of each phase (illustrated in Fig. 1), the resulting data set can be qualified as an unbalanced panel structure. The amount of data available for each household is, on average, about four years (Table 2).

To verify that our experiment allows reasonable conclusions from the available sample, we conducted a power analysis for multiple regression following Cohen's method (Cohen, 2013). At the power level of 0.8, a significance criterion of 0.1, a medium effect size ($F^2 = 0.15$) which can be seen typical for home energy audits (Delmas et al., 2013), and two independent variables (heating degree days, inspected), this analysis led us to a suggested sample size of 54 households, i.e. 27 households to be inspected.

We consider an α level of 0.1 as the threshold in evaluating statistical differences as appropriate for our analysis. That is because statistically significant differences across groups are particularly hard to obtain in campaigns targeting household electricity usage due to the large variance in consumption (Allcott, 2011). Besides that, an α level of 0.1 is acceptable for our investigated intervention that involves merely economic risks (Anderson et al., 2020).

3.2. Electricity consumption data

We prepared the electricity consumption data from both utilities in the same way. As many households were only equipped with a single meter, but others had multiple meters (e.g. one for the heat pump and one for the remaining household consumption).[3] We added the measurements from all meters of each household and aggregated them into monthly sums. Few households showed missing values, typically for a couple of hours to days. In this case, we excluded the entire month in which the measurement error occurred. In addition, we removed households where the consumption values were mixed with feed-in values from photovoltaic systems, those with obviously wrong documentation on heat pump existence, and such with unusually high or low consumption patterns (e.g. indicating commercial adjunct use or cottages). Finally, we excluded all households for which we did not receive data on an entire full-year heating period and those with less than three months of data in the treatment phase. This led to total consumption values for 14,815 months (Table 2).

Table 3 and the boxplots in Figure A.2 in the appendix show the

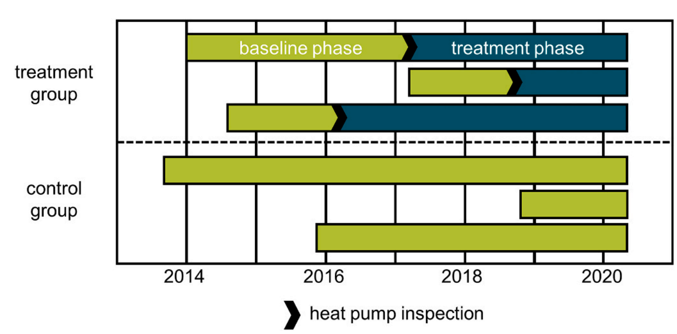

Fig. 1. Schematic illustration of the quasi-experiment.

[2] The main reason for the lack of data was the low penetration rate of smart meters in Supplier A's distribution area when the inspection service was introduced in 2015.

[3] A single metering point is typical for single-family houses that do not offset heat costs to tenants and if utilities do not offer a special tariff for heat pumps. Since each additionally installed meter is charged (in the EU between EUR 180 and EUR 200 installation cost plus operating expense (EC, 2014)), households have no advantage of installing several meters.

Table 2
Data per group and phase.

Data characteristics	Treatment *(Utility A)*	Control *(Utility B)*	Total
Households	41	256	297
Months of data	2,353	12,462	14,815
in baseline phase	1,134	12,462	–
in treatment phase	1,219	–	–
Average months per household	57.4	48.7	49.9
in baseline phase	27.7	48.7	–
in treatment phase	29.7	–	–

We calculated the consumption statistics (mean, median) for each household during the baseline phase for the entire period but separately for the heating phase (September–April) and the non-heating phase (May–August).

consumption characteristics for both groups, the mean value, the standard deviation (SD), and the range. As the treatment and the control group were from two different utilities in neighbouring regions, we ensured comparability of the consumption values in both groups by performing a Kruskal-Wallis rank sum test. The p-values in Table 3 show no statistically significant difference in consumption characteristics between the treatment and control groups during the baseline phase for all except one variable, but there the effect size is marginal ($\eta^2 = 0.01$).

3.3. Weather data

Outside temperature considerably influences the heat demand of buildings and thus the electricity consumption of heat pumps. To include the influence of weather in our analysis, we used the "TabsD" data set from MeteoSwiss (2017), which contains interpolated averaged air temperature values for each point in a 1 × 1 km grid over Switzerland. We calculated the heating degree days (HDD) per month for each grid point. HDD describes the sum of the daily differences between an assumed indoor temperature (we use 20 °C) and the average outdoor air temperature when it is below a threshold (we use 12 °C) for which the heating system typically operates. We matched HDD data to the monthly electricity consumption data based on the nearest grid point to each household and constructed an unbalanced panel data set.

4. Results and discussion

This section introduces the analysis methods used to answer RQ1–3 and presents and discusses our findings.

4.1. Saving of heat pump inspections

To answer RQ1, we analysed our longitudinal data set using a linear regression approach. The most common methods to analyse such panel data are pooled, fixed effects (FE), and random effect models. FE models have the advantage over pooled models in capturing unobserved individual-specific effects (e.g. building size or isolation standard). Appendix B details the model selection and suggests using a FE model for our data set. We thus estimated the average saving through heat pump inspections with a two-way FE panel data regression using ordinary least squares with the R package PLM (Croissant and Millo, 2008; Millo, 2017) and defined the linear panel model as

$$Consumption_{it} = \beta_1 Inspected_{it} + \beta_2 HDD_{it} + c_i + \lambda_t + u_{it} \tag{1a}$$

where $Consumption_{it}$ is the electricity consumption of a household i in the month t, $Inspected_{it}$ is a dummy variable indicating that a consultant inspected a heat pump, HDD_{it} corresponds to the heating degree days at a household i in a given month t, c_i is the unobserved individual fixed effect for each household, λ_t is the time fixed effect in a respective month, and u_{it} is the error term. We estimated a robust covariance matrix and clustered the observations by household with the White-Arellano method to account for the well-known identified problems heteroscedasticity and serial correlated standard errors in panel data models also found in our data set (Arellano, 1987; Wooldridge, 2010). Appendix B shows the corresponding statistical tests. Table 4 displays the estimates of model 1a which leads to a point estimate for savings from heat pump inspections of 53.5 kWh per month, respective 642 kWh per year for each household, or 5.3% saving for the average heating

Table 4
Overall treatment effect of heat pump inspections (model 1a).

Independent variables	Consumption [kWh/month]		
	Estimate	[CI]	S.E.
Inspected	−53.53	[−130.83; 23.78]	46.99
HDD	0.96	[0.75; 1.16]	0.12***
Observations			
Number of households analysed n	n = 297		
Range of months analysed T	T = 12–86		
Number of household months N	N = 14,815		
R^2/R^2 adjusted	0.01/−0.02		

Notes: ***, **, *, and . indicate statistical significance at 0.1%, 1%, 5%, and 10%. Robust standard errors clustered (S.E.) by households are estimated using the White-Arellano method (Arellano, 1987); the 90% confidence interval is shown in square brackets.

Table 3
Electricity consumption of treatment (T) and control (C) group during the baseline phase.

Electricity consumption during baseline phase in kWh per month		Control	Treatment	Total	Difference between C and T (p-value)
Median consumption	Mean (SD)	1,000.1 (521.9)	901.8 (469.9)	986.6 (515.4)	0.348
	Range	237.4–3,809.7	97.5–2,131.5	97.5–3,809.7	
Mean consumption	Mean (SD)	1,048.3 (522.3)	959.8 (517.0)	1,036.1 (521.6)	0.310
	Range	261.7–3,787.6	171.3–2,808.5	171.3–3,787.6	
Mean consumption in the heating phase	Mean (SD)	1,232.4 (611.9)	1,202.8 (718.3)	1,228.3 (626.4)	0.637
	Range	325.4–4,265.3	248.8–4,101.3	248.8–4,265.3	
Mean consumption in the non-heating phase	Mean (SD)	629.7 (377.5)	536.5 (275.8)	616.8 (366.1)	0.230
	Range	120.7–2,877.9	61.9–1,386.7	61.9–2,877.9	
Mean consumption in heating phase − Mean consumption in non-heating phase	Mean (SD)	602.7 (363.9)	666.3 (506.0)	611.5 (386.2)	0.701
	Range	−388.8–2,621.7	60.0–2,714.6	−388.8–2,714.6	
Mean consumption in non-heating phase/Mean consumption in heating phase	Mean (SD)	0.5 (0.1)	0.5 (0.2)	0.5 (0.1)	0.045
	Range	0.2–1.3	0.2–0.9	0.2–1.3	

degree day in this data set.[4] However, the treatment effect is not statistically significant, most likely due to the wide variation in effect sizes among the households, which calls for a subgroup analysis.

4.2. Subgroups with different saving potential

To examine differences in the overall treatment effect and answer RQ2, we explore subgroups of households that reduced their consumption considerably, households that reduced little, and those that did not reduce. For that, we calculated the treatment effect for each household separately and grouped households then according to the relative order of the individual saving estimates.

We estimated the individual saving effect through the heat pump inspections with a variable coefficient (VC) model (Hsiao, 2014) that allowed us to systematically vary the coefficients *Inspected* and *HDD* for each household. Since this model estimates consumption differences independently for each household, it cannot control for individual, time-fixed, and regression-to-the-mean effects. This made it necessary for us to introduce a virtual treatment phase for the control group, which allowed us to estimate the consumption differences over time also for the control group, and thus compare the actual savings of the treatment group with the differences of the control group (to which we consistently refer to as "savings"). To realise the virtual treatment phase (Fig. 2), we divided the available data points of the control group, leaving the first half as the baseline phase and using the second half as the virtual treatment phase since this split (Table 2) is similar to the intervention phase of the treatment group. Even after removing half of the control group data to a virtual treatment phase, the consumption values between both groups during the baseline phase remained comparable (see Table A.1 in the appendix).

Based on the actual treatment phase (treatment group) and the virtual treatment phase (control group) that are both coded in $Inspected_{it}$, we estimate the following VC model

$$Consumption_{it} = \beta_{0i} + \beta_{1i} Inspected_{it} + \beta_{2i} HDD_{it} + u_{it} \tag{1b}$$

The results show wide variation in the individual estimates between the baseline and the (virtual) treatment phase. The individual estimates (β_{1i}) for the 41 households that received treatment range from more than 15,000 kWh (−50%) reduction to almost 5,000 kWh (+75%) increase per year. The control group shows similar variations, meaning that some households have substantially reduced or increased their consumption by chance. Next, we used the order of the estimates for each household to discriminate between several groups with different savings potentials in the treatment group. 61% of the households reduced their consumption after the inspection (we label this group *Savers*), whereas the other 39% showed no savings or even increased consumption (*Non-Savers*). Given that some *Savers* showed only modest while others showed substantial decreases, we split this group and defined the upper half as *High-Savers* (29% of the treatment group) and the lower half as *Low-Savers* (32% of the treatment group). We applied the same logic to the control group and divided them accordingly into High-Savers (25%), Low-Savers (25%), and Non-Savers (50%).

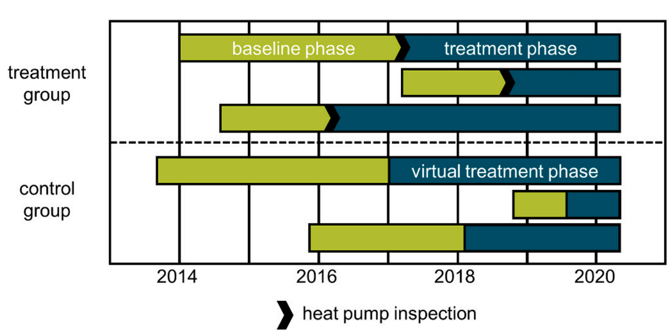

Fig. 2. Schematic illustration of the quasi-experiment adopted by a virtual treatment phase.

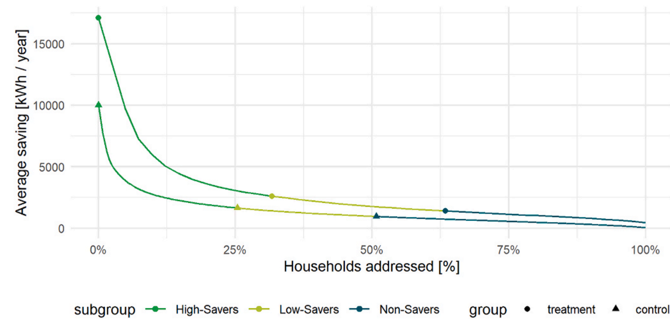

Fig. 3. Average saving per household depending on the percentage of households addressed.

We conclude that large differences in the saving potential among households lead to moderate effects without targeting. Conversely, we can expect high average savings when "high impact" households are identified upfront. We illustrate this situation in Fig. 3, where we depict the average annual savings for a varying share of households addressed under the hypothetical assumptions that energy efficiency service providers could identify ex-ante and sort households according to their savings. The upper line shows the average savings of the treatment group in this hypothetical campaign when starting with the High-Savers, followed by the Low-Savers and drops to almost zero when the Non-Savers are included. This logic also applies to the lower line that shows the savings of the control group (based on the virtual treatment phase) and is probably caused by the regression-to-the-mean effect, which also holds in parts for the treatment group. However, we can observe two aspects that indicate real savings for the treatment group. First, the average savings for the treatment group is larger for each share of customers addressed (e.g. 25%, 50%, and 100%) and thus indicates a larger actual saving. Second, the larger savings occur in the treatment group, although the control group has a slightly higher consumption in the baseline phase and thus a higher "saving" potential (see the consumption values for both groups with the new baseline for the control group in appendix Table A.1). The actual reduction of the treatment group through the intervention lies within both lines. Below, we demonstrate the extent to which such an ex-ante identification of High-Savers is possible using smart meter electricity data and estimating the actual saving potential.

4.3. Identifying households with saving potential

In an attempt to identify households that will generate large energy savings, we tested the power of twelve criteria that could help to find such "high impact" households (RQ3). For this, we used the previously defined groups of High-Savers, Low-Savers, and Non-Savers. We deliberately tested selection criteria available for many utilities, namely (1) consumption-based criteria from smart meter data, (2) characteristics of the heating system, and (3) the building: For the first category, we included consumption data at a higher aggregation level than the current smart metering infrastructure can measure and transmit to ensure that even companies with large reading intervals can implement such simple pre-selection routines. For the latter two categories, information on the heating system and the building is either publicly available or predictable based on smart meter data. For example, public cadastre databases contain information on ground source heat pumps, which often have to be registered for water protection reasons. In addition, the prediction of the heat pump type (Weigert et al., 2020) and the hot water

[4] The average heating degree day is 293.2 in our data set and is exemplarily given in the months October and April.

type (Hopf et al., 2018) was shown to be feasible based on smart meter data with a common temporal resolution of 15 minutes and weather data.

To evaluate the selection criteria, we first check for differences in each criterion between the considered subgroups. For that, we use a univariate statistic—Fisher's exact test for categorical variables and Kruskal-Wallis rank sum tests for continuous variables. We assessed differences between the treatment subgroups and summarised the results in Table 5. High-Savers and Low-Savers differ statistically significantly in four consumption characteristics measured before the intervention, namely in the *median consumption* ($p < 0.01$), the *mean consumption* ($p < 0.05$), the *mean consumption in the heating phase* ($p < 0.1$), and the *mean consumption in the non-heating phase* ($p < 0.05$). Particularly interesting is that the group of Non-Savers exhibited consistently lower electricity use before the intervention, which can have multiple reasons (e.g. already efficient installation, high energy awareness among users, low income). Moreover, the groups of Savers (High-Savers and Low-Savers combined) and Non-Savers differ statistically significantly in terms of the *type of the heat pump* ($p < 0.05$). To put that into perspective, Savers had installed a ground source heat pump almost twice as often as Non-Savers in our sample.

In Appendix C, we replicate this analysis with a more comprehensive multivariate approach that also includes the control group. This extended analysis confirms the median consumption and the heat pump type as significant pre-selectors. In addition, we found no evidence that being a Saver depends on the overall heat demand (e.g. given by the size of the building and the number of residents) or on other heating system characteristics (e.g. external water heating boilers, the existence of radiators or underfloor heating) than the heat pump type. These results are interesting because such characteristics were shown to be relevant to influencing the overall heat pump performance (Caird et al., 2012) but seem to be irrelevant for quick-wins induced by heat pump inspections. This multivariate analysis also revealed the difference between the heating and the non-heating phase as a possible selector, which was insignificant in the univariate analysis. We, therefore, decided to proceed with the impact investigation of the pre-selectors for the two variables, median consumption and heat pump type, since they indicate selective power in both analyses.

The second step in our evaluation of the selection criteria is an impact analysis. We estimate the average saving of a campaign that uses one selection criterion at a time using the same FE modelling approach that we introduced before to estimate the overall treatment effect without selection.

We, therefore, extend model 1a by adding pre-selectors as an independent variable in the form of dummies and interacting them with the treatment effect $Inspected_{it}$. This is done by binarising the pre-selector variable *median consumption* through a median split that cuts the groups into two halves, namely high consumers (HC) and low consumers (LC). In our notation, the dummy $Median^{HC} = 1$ denotes households whose median consumption is above the group's median (default value

Table 5
Subgroup analysis to identify selection criteria for household pre-selection.

Criterion	Savers			Non-Savers	p-value	
	Total (N = 25)	High-Savers (N = 12)	Low-Savers (N = 13)	Total (N = 16)	Saver vs Non-Savers	High-Savers vs Low-Savers
Median consumption in kWh per month					0.209	0.008
Mean (SD)	949.7 (418.0)	1,167.4 (459.4)	748.7 (254.5)	827.0 (547.2)		
Range	304.5–2,131.5	304.5–2,131.5	372.0–1,213.0	97.5–2,099.4		
Mean consumption in kWh per month					0.323	0.026
Mean (SD)	1,010.5 (512.4)	1,221.3 (605.0)	815.9 (322.7)	880.7 (530.8)		
Range	347.3–2,808.5	375.1–2,808.5	347.3–1,474.7	171.3–2,170.9		
Mean consumption in heating phase in kWh per month					0.336	0.057
Mean (SD)	1,282.8 (749.0)	1,537.1 (922.1)	1,048.0 (467.2)	1,077.8 (671.4)		
Range	394.0–4,101.3	429.6–4,101.3	394.0–1,978.6	248.8–2,664.8		
Mean consumption in non-heating phase in kWh per month					0.121	0.022
Mean (SD)	589.6 (280.2)	724.7 (293.4)	464.9 (207.4)	453.6 (255.5)		
Range	148.0–1,386.7	231.3–1,386.7	148.0–778.8	61.9–1,033.8		
Mean consumption in heating phase − Mean consumption in non-heating phase					0.789	0.253
Mean (SD)	693.2 (538.9)	812.4 (686.8)	583.1 (348.1)	624.2 (463.7)		
Range	60.0–2,714.6	81.2–2,714.6	60.0–1,352.9	101.7–1,630.9		
Mean consumption in non-heating phase/Mean consumption in heating phase					0.364	0.415
Mean (SD)	0.5 (0.2)	0.5 (0.2)	0.5 (0.2)	0.4 (0.2)		
Range	0.2–0.9	0.3–0.9	0.2–0.8	0.2–0.8		
Heated space area in m^2					0.134	0.604
Mean (SD)	252.2	256.7 (113.0)	248.2 (97.2)	215.1 (78.4)		
Range	80.0–560.0	80.0–560.0	135.0–525.0	126.0–400.0		
Number of residents					0.351	0.401
Mean (SD)	3.0 (1.6)	3.2 (1.5)	2.8 (1.6)	2.5 (1.0)		
Range	1.0–6.0	1.0–6.0	1.0–6.0	1.0–5.0		
Heat pump type					0.025	1.0
Ground	16 (64.0%)	8 (66.7%)	8 (61.5%)	4 (25.0%)		
Air	9 (36.0%)	4 (33.3%)	5 (38.5%)	12 (75.0%)		
Hot water type					0.522	0.695
Heating syst.	13 (52.0%)	7 (58.3%)	6 (46.2%)	6 (37.5%)		
Ext. boiler	12 (48.0%)	5 (41.7%)	7 (53.8%)	10 (62.5%)		
Floor heating					0.685	1.0
Non-existent	5 (20.0%)	2 (16.7%)	3 (23.1%)	2 (12.5%)		
Existent	20 (80.0%)	10 (83.3%)	10 (76.9%)	14 (87.5%)		
Radiators					1.0	0.226
Non-existent	16 (64.0%)	6 (50.0%)	10 (76.9%)	10 (62.5%)		
Existent	9 (36.0%)	6 (50.0%)	3 (23.1%)	6 (37.5%)		

Notes: The table shows the mean and standard deviation in parenthesis for continuous variables and the p-value of Kruskal-Wallis rank sum tests. For categorical variables, it shows the frequency of households for each category, the percentage in parenthesis, and the p-value of a Fisher's exact test.

of 0). Vice versa, the superscript “LC” stands for low consumers. In the case of the categorical pre-selector variable *HeatPump* type, the superscript “ground” stands for ground source heat pumps, while we use the type “air source” as the default case (default value of 0).

Before we gauge the impact of these pre-selection criteria, there is one important aspect left to consider. Given that we use the dependent variable *Consumption* to derive the pre-selection criteria $Median^{HC}$ that serves in the model also as independent variables, the treatment effect might be influenced by the regression-to-the-mean effect introduced before. We correct for this aspect as we derive the pre-selection criteria also for the control group using the previously introduced virtual treatment phase. By adding interaction terms in the models between the pre-selection criteria and the groups, we estimate the difference in consumption between the high consumers from the treatment and the control group and thus calculate the true difference in treatment without the regression-to-the-mean effect. This results in the following model specification

$$Consumption_{it} = \beta_1 Inspected_{it} + \beta_2 Inspected_{it} \times Control \times Median_i^{LC} + \beta_3 Inspected_{it} \times Treatment \times Median_i^{LC} + \beta_4 Inspected_{it} \times Treatment \times Median_i^{HC} + \beta_5 HDD_{it} + c_i + \lambda_t + u_{it} \tag{2}$$

$$Consumption_{it} = \beta_1 Inspected_{it} + \beta_2 Inspected_{it} \times HeatPump_i^{ground} + \beta_3 HDD_{it} + c_i + \lambda_t + u_{it} \tag{3}$$

The estimates of models 2–3 show a clear picture (Table 6). Both criteria have a statistically significant impact on consumption. In model 2, for example, households with an above-median consumption in the baseline phase save 152 kWh per month more (p-value = 0.097) or save 15.2% for the average heating degree day in this data set than households in the control group with the same selection criteria (the reference group in this model). This is not the case for the treated households with a low median baseline consumption which did not save compared to the part of the control group with a low baseline consumption. The point estimate suggests an increase of 34 kWh per month that is not statistically different from zero (p-value = 0.218). Performing a pre-selection based on the heat source of the heat pump (model 3) even leads to a treatment effect that is 239 kWh higher (p-value = 0.015), or 23.5% in this filtered data set, for ground source heat pumps compared to the not-selected group (air source heat pumps). Please note that a regression-to-the-mean correction is unnecessary for model 3 as the selection does not depend on the dependent variable. Both pre-selection criteria increased the average savings considerably compared to our first model (1a), which estimates overall savings of 54 kWh per month. Please note that the savings reported here are monthly averages, and annual savings are twelve times these figures, amounting to 1,805 kWh per year and household for a selection based on a high median consumption and 2,158 kWh for a selection based on the heat pump type. In the subsection that follows, we estimate the possible financial impact of an energy-saving campaign that uses the identified pre-selection criteria.

4.4. Using pre-selection criteria in efficiency campaigns

Efficiency measures, like the heat pump inspections we analyse in this study, are often not fully paid by customers who order the service but are co-financed by the public, making it necessary to determine the effectiveness of such measures in terms of saved energy, cost, and benefit.

Our study demonstrated that not all inspected households realised energy efficiency gains, but society bears a large share of their expenses. Therefore, the cost-efficiency of such programs and the purposeful pre-selection of candidates are crucial and can help to achieve both targets. Hence, we analyse the monetary effect of using the identified selection criteria in an analytics software solution to automatically pre-select high-impact households.

In doing so, we consider two exemplary campaigns: In campaign A, customers order the service, which means they select themselves for the campaign. In campaign B, utilities offer services only to the 50% of households for which they assume high savings, using one of the criteria described above. Table 7 shows the net benefit for both selection methods: We take the original cost to conduct a heat pump inspection, which was CHF 400 (USD 426) in our case. For campaign B, we assume that a smart meter infrastructure is already in place (e.g. mandated or to support the meter to cash process) but that an analysis system is required to process the data and compute the values of the selection criteria for all customers. We assume licence fees of CHF 10 (USD 10.7) for each analysed household for this system. In addition, we use the estimates from model 1a for campaign A and the one from model 2 for campaign B and assume an electricity price of 0.207 CHF/kWh (0.223 USD/kWh).[5]

Overall, the net benefit per household is approximately zero for campaign A when accumulating the efficiency gains over three years. In contrast, the pre-selection with an analytical tool in campaign B creates a net benefit of CHF 701 (USD 747). However, both types of campaigns can also be associated with higher risks or higher benefits. Thus, we use the 90% confidence interval of the estimates in Table 6 to express the upper and lower limit of the expected net benefit that we present in Fig. 4. The results show that the lower bound of the net benefit for both campaigns can be negative, with about CHF −577 (USD −615) for campaign A and CHF −633 (USD −675) for campaign B, which can be explained for B in parts by the licence fees for the analytical system. The upper bound of the net benefit for campaign A (CHF 575, USD 613) does not exceed the average net benefit of campaign B (CHF 701, USD 747), while B can yield up to CHF 2,034 (USD 2,168) in savings. From a risk/benefit perspective, we sum up that the risk for a negative net benefit is roughly similar for both campaigns. Still, campaign B’s average and upper limit net value suggest higher economic value.

5. Conclusion and policy implications

Heat pumps can extract energy from the ground, soil, or ambient air, turning their owners into operators of small sustainable power plants. However, they rely on electricity to operate, and the increasing prevalence of heat pumps is putting a strain on electricity grids. The efficiency of heat pump systems in operation is therefore of growing importance. Each installation that does not perform as desired wastes resources and causes unnecessary operating costs. Some of the causes for low performance are difficult to solve after installation. Still, energy consultants can easily remedy many issues with a single inspection visit.

Our study examined the impact of a campaign in which experts investigated residential heat pumps to remedy easily detectable and solvable issues (e.g. optimisation of the settings of the heating control unit, including user training). The analysis of longitudinal electricity consumption data of 41 treated households revealed that they saved on average 642 kWh (5.3%) per year compared to a control group of 256 non-treated households. However, the treatment effects varied widely and were overall not statistically significant: While a substantially large subgroup (61%) of the treatment group saved energy compared to their baseline consumption, there is also another subgroup of Non-Savers (39% of our sample) that increased their consumption. These results suggest that heat pump inspections yield substantial energy conservation only if households are purposefully pre-selected.

To carry out such a pre-selection, we evaluated several easy-to-obtain criteria for electric utilities (e.g. the median consumption and

[5] In 2020, the average electricity price for residential households in Switzerland was 0.207 CHF/kWh; see: https://www.elcom.admin.ch/elcom/de/home/dokumentation/medienmitteilungen.msg-id-76327.html, last accessed Sept 09, 2021).

Table 6
Analysis of the identified criteria to pre-select high potential heat pumps.

Independent variables	Consumption [kWh/month]					
	Model 2			Model 3		
	Estimate	[CI]	S.E.	Estimate	[CI]	S.E.
Inspected	1.73	[−26.43; 29.89]	17.12	59.16	[−37.92; 156.24]	59.02
HDD	0.95	[0.75; 1.15]	0.12***	1.35	[0.33; 2.36]	0.62*
Inspected × Control × Median^{LC}	12.68	[−14.94; 40.29]	16.79			
Inspected × Treatment × Median^{LC}	46.84	[−2.83; 96.52]	30.20			
Inspected × Treatment × Median^{HC}	−152.15	[−302.93; −1.37]	91.67.			
Inspected × HeatPump^{ground}				−239.02	[−401.07; −76.98]	98.52*
Observations						
Num. households analysed n		n = 297			n = 41	
Range of months analysed T		T = 12-86			T = 19-86	
Num. household months N		N = 14,815			N = 2,353	
R^2/R^2 adjusted		0.01/−0.01			0.03/−0.02	

Notes: ***, **, *, and . indicate statistical significance at 0.1%, 1%, 5%, and 10%. Robust standard errors clustered (S.E.) by households are estimated using the White-Arellano method (Arellano, 1987); the 90% confidence interval is shown in square brackets.

Table 7
Net benefit of heat pump inspection per household based on the customer selection method.

	Campaign A: Heat pump inspections by self-selection	Campaign B: Heat pump inspections by pre-selection (using an analytical system)
Cost per inspection	CHF 400 USD 426	CHF 400 USD 426
Licence fees to identify high-impact households per household analysed	CHF 0 USD 0	CHF 10 USD 11
Licence fees per high-impact household	CHF 0 USD 0	CHF 10/50% households selected = CHF 20 USD 21
Annual savings in kWh per household through heat pump inspection	53.53 kWh × 36 months = 1,927 kWh	150.42 kWh × 36 months = 5,415 kWh
Electricity price per kWh	0.207 CHF/kWh 0.221 USD/kWh	0.207 CHF/kWh 0.221 USD/kWh
Net benefit per household	1,927 kWh × 0.207 CHF/kWh – CHF 400 = **CHF −1** USD −1	5,415 kWh × 0.207 CHF/kWh – CHF 400–CHF 20 = **CHF 701** USD 747

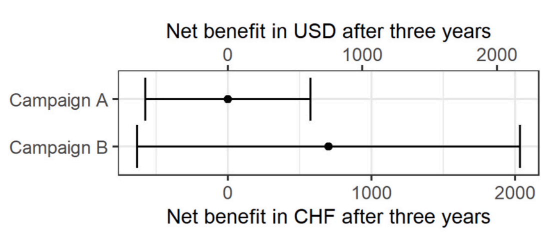

Fig. 4. Net benefit of heat pump inspection per household, including 90% confidence interval.

the heat pump type) in this study. With these criteria, the cost-effectiveness of a campaign can be increased considerably. According to our assessment, a campaign that uses the suggested pre-selection criteria "median consumption" can lead to savings of 1,805 kWh (15.2%) per year and household, and even to 2,158 kWh (23.5%) saving if the heat pump source is known. In such a campaign, the more conservative estimate yields a three-year net benefit of CHF 701 (USD 747) compared to a households' self-selection campaign, which is, on average, not economically viable CHF −1 (USD −1). Our findings have implications for energy policy and utility companies, which we describe below. We list the limitations of our study after that.

5.1. Implications for energy policy

A sustainable and efficient energy supply has been at the top of the political agenda not only since the Paris Agreement. In this context, the electrification of heat generation and the efficiency of heating systems become more critical. Our study shows the high savings potential of currently installed heat pumps—which will most likely remain in operation for more than a decade—and confirms findings of earlier studies that many heat pumps in the field perform lower than expected (Caird et al., 2012; Chesser et al., 2021; Puttagunta et al., 2010; Qiao et al., 2020; Yin et al., 2019). While policy measures improving heat pump performance have so far dominantly focused on subsidising new installations of certain efficiency standards (EC, 2013a, 2013b, 2009; EP, 2017), our study results urge to focus on the existing stock of installations. *A powerful instrument to increase the energy efficiency of installed heat pumps are on-site visits by experts that identify easy-to-fix issues and provide user training.* This measure demonstrates significant net benefits *if households are purposefully pre-selected*, as demonstrated in this paper. If such inspections are offered to all customers, the effect per inspection (and thus the cost efficiency) drops considerably, as the savings effects vary greatly depending on the household.

Deciding which households to include in a campaign requires a data-driven assessment of their potential. For this, we propose a set of selection criteria that utility companies can derive using electricity consumption data recorded by common smart meters. The investigated criteria are not limited to heat pump installations that are measured separately but are based on the aggregated electricity consumption data of an entire household. This makes utilities and smart meter infrastructure operators the key players in this task. However, these *organisations need to be legally equipped to access and process such data for the purpose of energy efficiency interventions*. In many countries, the use of smart meter data is limited to a small number of use cases, often for unbundling reasons which should prevent players from using data obtained under a monopoly for use cases outside a narrowly defined scope. In Switzerland, for example, regulations prevent utilities from using smart meter data derived from grid operations when this could potentially lead to a competitive advantage over energy consultants. Moreover, households are often unaware that using their smart meter data could lead to significant energy and money-saving and thus avoid giving user consent explicitly for data processing. Thus, policymakers need to strike a balance between society's need for competition, data protection, and sustainability goals.

Even when legal hurdles to process data for energy efficiency interventions are removed or when users provide their informed consent for data processing, technical issues hinder data sharing with service providers. One of these hurdles is the lack of interfaces that make metering data available to external service providers, even if households provided consent. *Smart grid operators should give the customers interfaces to exchange data with service providers that are easy to use, safe, and well documented.* In fact, policymakers should pave the way for the use of available data to detect poorly adjusted heating systems.

Furthermore, we see great potential in using additional (open) data sources. For example, cadastre data for water or noise protection document characteristics on installed heat pumps (e.g. place, date of installation, and type). Such data could enhance analytical systems to support energy efficiency services. To date, such data is only publicly available for a few regions and is often distributed among different authorities (e.g. regional departments of planning and building inspection). Governmental organisations that possess such data, thus, should integrate their data pools, and offer a uniform, central data interface to make them available for research and service companies.

5.2. Implications for utilities

Utilities are often required to implement public mandates to increase energy efficiency (Alberini and Towe, 2015; Cho et al., 2019; EU, 2021; Taylor et al., 2014). Our study shows that heat pump inspections are measures that can achieve efficiency targets at desirable costs—provided that the targeting of households is based on appropriate pre-selection. Smart meter data make such pre-selection possible. Thus, utilities can recoup their heavy investments in smart metering infrastructure, which are only slowly amortising so far. Furthermore, implementing such analytical solutions can foster their data analytics capabilities and value creation mechanisms (Hopf et al., 2022). On the other side, customers can benefit from targeting, as heat pump inspections rely on time investment (and often also a financial one). With a pre-selection campaign, their effort-/cost-to-benefit ratio would, on average, improve significantly. At the same time, the risk of a negative net benefit for customers stays comparable to customers in a self-selection campaign. The risk of ordering a service that does not pay off can be reduced for the customer with an aligned cost-sharing and benefit model, where the utility bears the cost but benefits together with the household from the improvements. Such models can support households suffering from energy poverty (Guertler, 2012). Last but not least, utilities must set the legal basis for such an approach and ask their customers to process their metering data to use them in targeted campaigns. The results of our research provide a legitimate reason to do so.

5.3. Limitations and future work

Our study is one of the first that examines the impact of heat pump inspections on the electricity consumption of residential homes in the field using a longitudinal dataset involving real consumer behaviour. Through its exploratory nature, our study comes with several limitations that, at the same time, pave the way for future research works.

We used the opportunity to examine heat pump inspections that a utility company had already conducted; our influence on the study design was limited. For example, we had limited information about study participants, dwelling attributes, etc., for both the treatment and control groups. In addition, our study includes smaller residential houses. We, therefore, motivate future research to validate our findings with advanced study designs that control for potentially confounding factors, sample biases, and influencing factors on the consumers' willingness to participate in heat pump inspection campaigns, and they could run additional randomisation checks between treatment and control groups. Such studies could also include larger heat pump installations (e.g. multi-family buildings or businesses) and could use more detailed electricity consumption data (e.g. separately measured heat pump and household consumption or increased measurement intervals). Furthermore, it is conceivable that our study design which used the overall electricity consumption instead of the pure heat pump consumption, led to a more conservative saving estimate. This is because the use of other appliances (i.e. entertainment electronics) also correlates with the time of the year (i.e. winter) and thus with the heating degree days used for control in our analysis.

The efficiency consultants not only remedied easy-to-fix measures during the heat pump inspection but also trained users and listed recommendations for future improvements (e.g. installing a speed-regulated circulating pump). It remains unclear to what extent the inspection only led to immediate technical adjustments or initiated residents' future actions or behaviour changes, such as implementing further adjustments to the heating system and influencing their behaviour when using other electricity-consuming appliances. However, the combination of technical solutions and user training is often given in such initiatives, so we believe it is helpful to examine the initiative as a whole. Which part of the effects result from long-lasting measures and which stem from behaviour changes that might not be persistent is also an open question; however, the observation period in our study was relatively long, with about four years of data per household, so we consider this limitation to be minor. Yet, future research should investigate the longitudinal persistency of such interventions and potential spill-over effects to other domains.

CRediT authorship contribution statement

Andreas Weigert: Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration. **Konstantin Hopf:** Methodology, Data curation, Writing – review & editing, Supervision, Funding acquisition. **Sebastian A. Günther:** Methodology, Software, Validation, Formal analysis. **Thorsten Staake:** Conceptualization, Methodology, Resources, Writing – review & editing, Visualization, Supervision, Funding acquisition.

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.

Acknowledgements

This work was supported by the Swiss Federal Office of Energy [grant number SI/501521-01]; and the German Federal Ministry for Economic Affairs and Energy [grant number 0350010].

The authors would like to thank Lorenz Deppeler, Marina Gonzalez Vaya, Hardy Schröder, and Michael Peter from the utility "Elektrizitätswerke des Kantons Zürich" (EKZ) for providing heat pump inspection and electricity consumption data and as well as fruitful discussions and insights on typical heat pump problems, André Rast from the utility "Centralschweizerische Kraftwerke AG" for providing electricity consumption data, and Peter Molnar from the ETH Zurich for providing weather data. Finally, the authors want to thank Kathrin Schmitt and Leonard Michels for their valuable comments on the early versions of this manuscript.

APPENDIX A. Details on the sample

Two additional analyses ensure the comparability of the treatment and the control group before and after introducing the virtual treatment phase and enriches, therefore, the similarity checks between both groups that we first introduced in Section 3.2. Table A.1 shows that the groups do not differ significantly (Kruskal-Wallis rank sum tests) in their consumption statistics. Figure A.2 shows boxplots illustrating their similarity visually.

For the control group, we used the full data set for the first analyses and half of the data set after we introduced the virtual treatment phase (we defined the virtual treatment phase as the second half of the initial baseline phase). The data of the treatment group remained unchanged. The p-values in Table A.1 show no statistically significant difference in consumption characteristics between the treatment and control with full and half data during the baseline phase. Only one variable—heating phase/non-heating phase—is different in the case of the full data set. Since only the ratio but not the difference between the heating and non-heating phases is statistically different, we do not think this points to a substantial difference overall, given that the effect size is small ($\eta^2 = 0.01$).

Table A.1
Electricity consumption during baseline phase for the treatment (T) and control group, with full data (C_{full}) and half data (C_{half})

Electricity consumption during baseline phase in kWh per month		T	C_{full}	C_{half}	p-value, difference between	
					T and C_{full}	T and C_{half}
Median consumption	Mean (SD)	901.8 (469.9)	1,000.1 (521.9)	943.7 (525.2)	0.348	0.947
	Range	97.5–2,131.5	237.4–3,809.7	178.9–3,815.6		
Mean consumption	Mean (SD)	959.8 (517.0)	1,048.3 (522.3)	1,018.7 (537.5)	0.310	0.633
	Range	171.3–2,808.5	261.7–3,787.6	249.9–3,908.7		
Mean consumption in the heating phase	Mean (SD)	1,202.8 (718.3)	1,232.4 (611.9)	1,241.1 (643.8)	0.637	0.679
	Range	248.8–4,101.3	325.4–4,265.3	308.3–4,286.5		
Mean consumption in the non-heating phase	Mean (SD)	536.5 (275.8)	629.7 (377.5)	621.7 (390.4)	0.230	0.357
	Range	61.9–1,386.7	120.7–2,877.9	90.2–3,135.9		
Mean consumption in heating phase – Mean consumption in non-heating phase	Mean (SD)	666.3 (506.0)	602.7 (363.9)	625.6 (424.3)	0.701	0.742
	Range	60.0–2,714.6	−388.8–2,621.7	−350.9–3,161.2		
Mean consumption in non-heating phase/Mean consumption in heating phase	Mean (SD)	0.5 (0.2)	0.5 (0.1)	0.5 (0.2)	0.045	0.127
	Range	0.2–0.9	0.2–1.3	0.1–1.2		

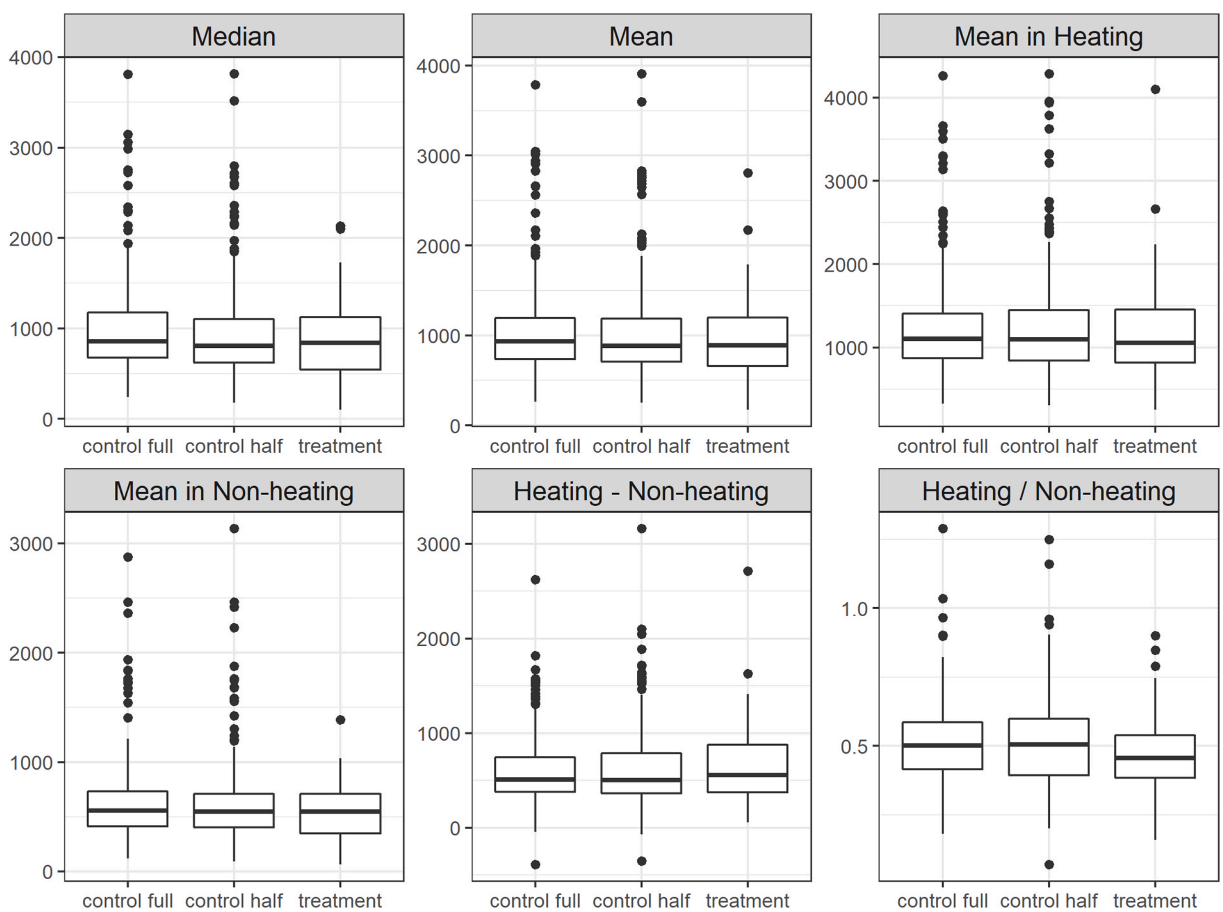

Figure A.2. Boxplots for electricity consumption of treatment and control group during the baseline phase

APPENDIX B. Details on the model selection

Further statistical tests helped us to choose an appropriate model to estimate the overall saving potential that we presented in Section 4.1. We performed two statistical tests to decide which type of model (pooled, FE, or random effects) is appropriate. First, we conducted an F-Test to compare a pooled and FE model and can reject H_0 that all fixed effects are equal to zero $\mathrm{F}(455,\ 14749)\ =\ 132.95,\ \mathrm{p}\ < 0.001$. Second, we show with the Hausman test that the FE model should be used rather than the random effects model $\mathrm{X}^2(2) = 42.417,\ \mathrm{p} < 0.001$ (Hausman, 1978).

The standard errors of the FE model show both heteroskedasticity and serial correlation. The former is discovered by applying the Breusch-Pagan test $\mathrm{BP}(2)\ = 115.39, \mathrm{p} < 0.001$, the latter by applying the Breusch-Godfrey test $\mathrm{X}^2(3)\ =\ 8464.1,\ \mathrm{p}\ <\ 0.001$ (Breusch, 1978; Breusch and Pagan, 1979).

APPENDIX C. Details on the pre-selection criteria

We performed two additional multivariate analyses to identify pre-selection criteria for households with saving potential and thereby confirmed the results of the univariate analyses in Section 4.2. We apply logistic regression to estimate log odds for the proposed selection criteria (see Table 5) and control thereby for multivariate and regression-to-the mean effects. For that, we use the two dependent variables (Savers vs Non-Savers, High-Savers vs Low-Savers) that we derived before by ranking the study participants according to the estimates from the VC model.

First, we estimate in model C.M1 the influence of all available selection criteria on the class Saver (vs being a Non-Saver). To be in line with the analysis in the main article, we consider that a utility has basic information on the building, the heating system, and access to smart meter data, and thus view all variables that are available for the treatment group. In the regression analysis, we must resolve strong correlations between the variables since they would lead to multicollinearity in the logistic regression model. The affected variables were the mean baseline consumption in the heating phase (***heating***) and the mean baseline consumption in the non-heating phase (***non-heating***). In addition, both variables correlate strongly with the overall baseline consumption (***mean*** and ***median***). Therefore, we used the difference between the heating and non-heating phase as a variable and removed other strongly correlating variables. The logistic regression model is, therefore

$$P(y_1 = 1|x) = \beta_0 + \beta_1\ Median + \beta_2(Heating - NonHeating) + \beta_3 Heated\ space\ area + \beta_4 Number\ of\ residents + \beta_5 Heat\ pump\ type + \beta_6 \text{Hot water type} + \beta_7 \text{Floor heating}\ + \beta_8 \text{Radiators} +\ u \quad (\text{C.M1})$$

where y_1 represents the first dependent variable and is defined as follows:

$$y_1 = \begin{cases} 0\ if\ \ a\ household\ is\ a\ NonSaver \\ 1\ if\ \ a\ households\ is\ a\ Saver \end{cases}$$

The results in Table C.1 show that only the ground source heat pump type significantly increases the log odds of being a Saver. We could not find support for other variables that describe a characteristic of the heating system (hot water type, floor heating, and radiators) or building (number of residents, heated space area). Therefore, the results from the multivariate approach are in line with the univariate approach in the main article and confirm the influence of the heat pump type that seems to be independent of other characteristics under study.

In the second analysis, we are interested in identifying High-Savers and assume that only characteristics on the baseline consumption, and thus on pure smart meter data, are available. Therefore, we can include the control group in model C.M2 (based on the introduced virtual treatment phase) and control for the regression-to-the-mean effect by adding interaction terms between selection criteria. We formulate

$$P(y_2 = 1|x) = \beta_0 + \beta_1\ Median + \beta_2(Heating - NonHeating) + \beta_3\ Median\ \times GroupTreatment + \beta_4(Heating - NonHeating)\ \times GroupTreatment +\ u \quad (\text{C.M2})$$

where y_2 represents the second dependent variable and is defined as follows:

$$y_2 = \begin{cases} 0\ if\ \ a\ household\ is\ a\ LowSaver \\ 1\ if\ \ a\ households\ is\ a\ HighSaver \end{cases}$$

The results in Table C.1 show that two variables significantly influence the log odds of being a High-Saver. The median baseline consumption (interacted with the treatment group) positively affects the High-Saver's log odds. On the other hand, the difference between the consumption in the heating phase and the non-heating phase (which are also interacted with the treatment group) decreases the log odds of being a High-Savers. This analysis confirms the results of the univariate analysis. We found again support for the median consumption as pre-selection criteria for High-Savers. This effect is stable when we control for the regression-to-the-mean effect.

This additional analysis further emphasises that a utility should pre-select households with high median consumption. Finally, the results indicate that the treatment group does not increase the log odds of being a High-Saver. This underlines the importance of pre-selection criteria to identify promising households for a campaign since the self-selection seems not to work on average, at least in our sample.

Table C.1
Subgroup analysis to identify selection criteria for household pre-selection

Independent variables	Model C.M1				Model C.M2			
	Saver vs Non-Saver				High-Saver vs Low-Saver			
	Estimate	[CI]	S.E.	p-value	Estimate	[CI]	S.E.	p-value
(Intercept)	−0.458	[−3.506; 2.590]	1.853		−2.229	[−3.167; −1.291]	0.570	***
Median	−0.000	[−0.003; 0.002]	0.002		0.002	[0.001; 0.003]	0.001	**
(Heating-Non Heating)	0.000	[−0.002; 0.002]	0.001		0.001	[−0.000; 0.002]	0.001	
Heated space area	0.002	[−0.007; 0.010]	0.005					
Number of residents	0.344	[−0.245; 0.933]	0.358					
Heat pump type$_{\text{ground}}$	1.762	[0.240; 3.285]	0.925	.				
Hot water type$_{\text{External Boiler}}$	0.044	[−1.268; 1.357]	0.798					
Floor heating$_{\text{Existent}}$	−0.858	[−2.892; 1.176]	1.237					
Radiators$_{\text{Existent}}$	−0.803	[−2.378; 0.772]	0.957					
Group$_{\text{Treatment}}$					−2.643	[−6.228; 0.942]	2.179	
Group$_{\text{Treatment}}$ × Median					0.006	[0.000; 0.011]	0.003	.
Group$_{\text{Treatment}}$ × (Heating–Non Heating)					−0.004	[−0.008; −0.001]	0.002	*
AIC	64.312				181.983			
BIC	79.734				199.926			
Log Likelihood	−23.166				−84.992			
Deviance	46.312				169.983			
Num. obs.	41				147			

Notes: ***, **, *, and . indicate statistical significance at 0.1%, 1%, 5%, and 10%.

References

Adams, P.W.R., Mezzullo, W.G., McManus, M.C., 2015. Biomass sustainability criteria: greenhouse gas accounting issues for biogas and biomethane facilities. Energy Pol. 87, 95–109. https://doi.org/10.1016/j.enpol.2015.08.031.

Alberini, A., Towe, C., 2015. Information v. energy efficiency incentives: evidence from residential electricity consumption in Maryland. Energy Econ. 52, S30–S40. https://doi.org/10.1016/j.eneco.2015.08.013.

Allcott, H., 2011. Social norms and energy conservation. J. Public Econ., Special Issue: The Role of Firms in Tax Systems 95, 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003.

Anderson, B., Rushby, T., Bahaj, A., James, P., 2020. Ensuring statistics have power: guidance for designing, reporting and acting on electricity demand reduction and behaviour change programs. Energy Res. Social Sci. 59, 101260 https://doi.org/10.1016/j.erss.2019.101260.

Arellano, M., 1987. Computing robust standard errors for within-groups estimators. Oxf. Bull. Econ. Stat. 49, 431–434.

Barbetta, G.P., Canino, P., Cima, S., 2015. The impact of energy audits on energy efficiency investment of public owners. Evidence from Italy. Energy 93, 1199–1209. https://doi.org/10.1016/j.energy.2015.09.117.

Barnett, A.G., 2004. Regression to the mean: what it is and how to deal with it. Int. J. Epidemiol. 34, 215–220. https://doi.org/10.1093/ije/dyh299.

Bode, G., Thul, S., Baranski, M., Müller, D., 2020. Real-world application of machine-learning-based fault detection trained with experimental data. Energy 198, 117323. https://doi.org/10.1016/j.energy.2020.117323.

Breisig, D.V., Deutsch, N., Homann, J., Linden, C., Neuhaus, J., 2020. Opportunities and risks for Germany's heating industry in a competitive global environment. PwC.

Breusch, T.S., 1978. Testing for autocorrelation in dynamic linear models*. Aust. Econ. Pap. 17, 334–355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x.

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica 47, 1287–1294. https://doi.org/10.2307/1911963.

Broberg, T., Egüez, A., Kažukauskas, A., 2019. Effects of energy performance certificates on investment: a quasi-natural experiment approach. Energy Econ. 84, 104480 https://doi.org/10.1016/j.eneco.2019.104480.

Caird, S., Roy, R., Potter, S., 2012. Domestic heat pumps in the UK: user behaviour, satisfaction and performance. Energy Effic. 5, 283–301.

Chesser, M., Lyons, P., O'Reilly, P., Carroll, P., 2021. Air source heat pump in-situ performance. Energy Build. 251, 111365 https://doi.org/10.1016/j.enbuild.2021.111365.

Cho, H., Freyre, A., Bürer, M., Patel, M.K., 2019. Comparative analysis of customer-funded energy efficiency programs in the United States and Switzerland–Cost-effectiveness and discussion of operational practices. Energy Pol. 135, 111010 https://doi.org/10.1016/j.enpol.2019.111010.

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, 0 ed. Routledge. https://doi.org/10.4324/9780203771587.

Connolly, D., 2017. Heat Roadmap Europe: quantitative comparison between the electricity, heating, and cooling sectors for different European countries. Energy 139, 580–593. https://doi.org/10.1016/j.energy.2017.07.037.

Croissant, Y., Millo, G., 2008. Panel data econometrics in R: the plm package. J. Stat. Software 27, 1–43. https://doi.org/10.18637/jss.v027.i02.

Decuypere, R., Robaeyst, B., Hudders, L., Baccarne, B., Van de Sompel, D., 2022. Transitioning to energy efficient housing: drivers and barriers of intermediaries in heat pump technology. Energy Pol. 161, 112709 https://doi.org/10.1016/j.enpol.2021.112709.

Delmas, M.A., Fischlein, M., Asensio, O.I., 2013. Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975 to 2012. Energy Pol. 61, 729–739. https://doi.org/10.1016/j.enpol.2013.05.109.

Deng, J., He, S., Wei, Q., Li, J., Liu, H., Zhang, Z., Zhang, H., 2020. Field test and optimization of heat pumps and water distribution systems in medium-depth geothermal heat pump systems. Energy Build. 209, 109724 https://doi.org/10.1016/j.enbuild.2019.109724.

Domanski, P.A., Henderson, H.I., Payne, W.V., 2014. Sensitivity analysis of installation faults on heat pump performance (No. NIST TN 1848). Nat. Ins. Stand. Technol. https://doi.org/10.6028/NIST.TN.1848.

EC, 2014. Smart Grids and Meters [WWW Document]. Energy - Eur. Comm. URL https://ec.europa.eu/energy/topics/markets-and-consumers/smart-grids-and-meters_en#smart-grids-development.

EC, 2013a. Commission Regulation (EU) No 813/2013, OJ L.

EC, 2013b. Commission Delegated Regulation (EU) No 811/2013, OJ L.

EC, 2009. Directive 2009/125/EC, OJ L.

Ecoplan, 2015. Smart Metering Roll Out – Kosten und Nutzen. Bern. Ecoplan. https://www.ecoplan.ch/download/smmu_sb_de.pdf.

Eggimann, S., Hall, J.W., Eyre, N., 2019. A high-resolution spatio-temporal energy demand simulation to explore the potential of heating demand side management with large-scale heat pump diffusion. Appl. Energy 236, 997–1010. https://doi.org/10.1016/j.apenergy.2018.12.052.

EHPA, 2019. The European Heat Pump Market and Statistics Report. European Heat Pump Association, Brussels. https://www.ehpa.org/fileadmin/red/07._Market_Data/2019/2019_Executive_Summary.pdf.

energieschweiz, 2015, Finanzierung von Massnahmen zur Förderung der Energieeffizienz und der erneuerbaren Energien beim Strom, https://www.local-energy.swiss/dam/jcr:f65ecb19-ee70-47a5-bf6f-8c86846d9cc4/Finanzierung_von_Massnahmen_der_Energiestrategie.pdf.

Eom, Y.H., Yoo, J.W., Hong, S.B., Kim, M.S., 2019. Refrigerant charge fault detection method of air source heat pump system using convolutional neural network for energy saving. Energy 187, 115877. https://doi.org/10.1016/j.energy.2019.115877.

EP, 2017. REGULATION (EU) 2017/1369, OJ L.

EU, 2021. DIRECTIVE 2012/27/EU of the EUROPEAN PARLIAMENT and of the COUNCIL of 25 October 2012 on Energy Efficiency, Amending Directives 2009/125/EC and 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC [WWW Document]. URL https://eur-lex.europa.eu/eli/dir/2012/27/2021-01-01.

EU, 2018. Directive (EU) 2018/2002 of the European Parliament and of the Council of 11 December 2018 Amending Directive 2012/27/EU on Energy Efficiency (Text with EEA relevance.), vol. 328.

EU, 2012. Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency, Amending Directives 2009/125/EC and 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC Text with EEA Relevance.

Fei, H., Kim, Y., Sahu, S., Naphade, M., Mamidipalli, S.K., Hutchinson, J., 2013. Heat pump detection from coarse grained smart meter data with positive and unlabeled learning. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13. Presented at the 19th ACM SIGKDD International Conference. ACM Press, Chicago, Illinois, p. 1330. https://doi.org/10.1145/2487575.2488203. USA.

Fischer, D., Madani, H., 2017. On heat pumps in smart grids: a review. Renew. Sustain. Energy Rev. 70, 342–357. https://doi.org/10.1016/j.rser.2016.11.182.

Gleeson, C.P., Lowe, R., 2013. Meta-analysis of European heat pump field trial efficiencies. Energy Build. 66, 637–647. https://doi.org/10.1016/j.enbuild.2013.07.064.

Gottwalt, S., Ketter, W., Block, C., Collins, J., Weinhardt, C., 2011. Demand side management—a simulation of household behavior under variable prices. Energy Pol. 39, 8163–8174. https://doi.org/10.1016/j.enpol.2011.10.016.

Guertler, P., 2012. Can the Green Deal be fair too? Exploring new possibilities for alleviating fuel poverty. Energy Pol. 49, 91–97. https://doi.org/10.1016/j.enpol.2011.11.059.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271. https://doi.org/10.2307/1913827.

Hillberg, E., Zegers, A., Herndler, B., Wong, S., Pompee, J., Bourmaud, Jean-Yves, Lehnhoff, S., Migliavacca, G., Uhlen, K., Oleinikova, I., Phil, H., Norström, M., Persson, M., Rossi, J., Beccuti, G., 2019. Flexibility Needs in the Future Power System. https://doi.org/10.13140/RG.2.2.22580.71047.

Hopf, K., Sodenkamp, M., Staake, T., 2018. Enhancing energy efficiency in the residential sector with smart meter data analytics. Electron. Mark. 28, 453–473. https://doi.org/10.1007/s12525-018-0290-9.

Hopf, K., Weigert, A., Staake, T., 2022. Value creation from analytics with limited data: a case study on the retailing of durable consumer goods. J. Decis. Syst. https://doi.org/10.1080/12460125.2022.2059172.

Hsiao, C., 2014. Analysis of Panel Data, third ed. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139839327.

IEA, 2021. Renewables 2021. Paris.

IEA, 2020. Heat Pumps [WWW Document]. IEA. URL https://www.iea.org/reports/heat-pumps.

IEA, 2019. Renewables 2019: Analysis and Forecasts to 2024. Paris.

Jack, M.W., Mirfin, A., Anderson, B., 2021. The role of highly energy-efficient dwellings in enabling 100% renewable electricity. Energy Pol. 158, 112565 https://doi.org/10.1016/j.enpol.2021.112565.

Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D., Ciais, P., 2021. Global Assessment of Oil and Gas Methane Ultra-emitters (Preprint). Atmospheric Sciences. https://doi.org/10.31223/X5NS54.

Long, J., Xia, K., Zhong, H., Lu, H., Y, A., 2021. Study on energy-saving operation of a combined heating system of solar hot water and air source heat pump. Energy Convers. Manag 229, 113624. https://doi.org/10.1016/j.enconman.2020.113624.

Love, J., Smith, A.Z.P., Watson, S., Oikonomou, E., Summerfield, A., Gleeson, C., Biddulph, P., Chiu, L.F., Wingfield, J., Martin, C., Stone, A., Lowe, R., 2017. The addition of heat pump electricity load profiles to GB electricity demand: evidence from a heat pump field trial. Appl. Energy 204, 332–342. https://doi.org/10.1016/j.apenergy.2017.07.026.

Madani, H., Roccatello, E., 2014. A comprehensive study on the important faults in heat pump system during the warranty period. Int. J. Refrig 48, 19–25. https://doi.org/10.1016/j.ijrefrig.2014.08.007.

Marinelli, S., Lolli, F., Gamberini, R., Rimini, B., 2019. Life Cycle Thinking (LCT) applied to residential heat pump systems: a critical review. Energy Build. 185, 210–223. https://doi.org/10.1016/j.enbuild.2018.12.035.

MeteoSwiss (Ed.), 2017. Documentation of MeteoSwiss Grid-Data Products Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD.

Miara, M., Grünther, D., Langner, R., Helmling, S., Wapler, J., 2017. 10 Years of Heat Pumps Monitoring in Germany. Outcomes of Several Monitoring Campaigns. From Low-Energy Houses to Un-retrofitted Single-Family Dwellings. Presented at the 12th IEA Heat Pump Conf., Rotterdam.

Miara, M., Günther, D., Kramer, T., Oltersdorf, T., Wapler, J., 2011. Wärmepumpen Effizienz-Messtechnische Untersuchung von Wärmepumpenanlagen zur Analyse und Bewertung der Effizienz im realen Betrieb. Fraunhofer ISE Freibg. Ger.

Millo, G., 2017. Robust standard error estimators for panel models: a unifying approach. J. Stat. Software 82, 1–27. https://doi.org/10.18637/jss.v082.i03.

Murphy, L., 2014. The influence of energy audits on the energy efficiency investments of private owner-occupied households in The Netherlands. Energy Pol. 65, 398–407. https://doi.org/10.1016/j.enpol.2013.10.016.

Pater, S., 2019. Field measurements and energy performance analysis of renewable energy source devices in a heating and cooling system in a residential building in southern Poland. Energy Build. 199, 115–125. https://doi.org/10.1016/j.enbuild.2019.06.057.

Primc, K., Slabe-Erker, R., Majcen, B., 2019. Constructing energy poverty profiles for an effective energy policy. Energy Pol. 128, 727–734. https://doi.org/10.1016/j.enpol.2019.01.059.

Puttagunta, S., Aldrich, R.A., Owens, D., Mantha, P., 2010. Residential ground-source heat pumps: in-field system performance and energy modeling. GRC Trans. 34, 941–948.

Qiao, Z., Long, T., Li, W., Zeng, L., Li, Y., Lu, J., Cheng, Y., Xie, L., Yang, L., 2020. Performance assessment of ground-source heat pumps (GSHPs) in the Southwestern and Northwestern China: in situ measurement. Renew. Energy 153, 214–227. https://doi.org/10.1016/j.renene.2020.02.024.

Renz, S., 2020. Heat pump market development in Switzerland. IEA HPT Mag 39, 4.

Taylor, N.W., Jones, P.H., Kipp, M.J., 2014. Targeting utility customers to improve energy savings from conservation and efficiency programs. Appl. Energy 115, 25–36. https://doi.org/10.1016/j.apenergy.2013.10.012.

The Economist, 2021. What Is behind Rocketing Natural-Gas Prices? the Economist.

Thomaßen, G., Kavvadias, K., Jiménez Navarro, J.P., 2021. The decarbonisation of the EU heating sector through electrification: a parametric analysis. Energy Pol. 148, 111929 https://doi.org/10.1016/j.enpol.2020.111929.

Tounquet, F., Alaton, C., 2020. Benchmarking Smart Metering Deployment in the EU-28. final report, Publications Office, LU.

Ürge-Vorsatz, D., Cabeza, L.F., Serrano, S., Barreneche, C., Petrichenko, K., 2015. Heating and cooling energy trends and drivers in buildings. Renew. Sustain. Energy Rev. 41, 85–98. https://doi.org/10.1016/j.rser.2014.08.039.

Verkooijen, K.T., Stok, F.M., Mollen, S., 2015. The power of regression to the mean: a social norm study revisited. Eur. J. Soc. Psychol. 45, 417–425. https://doi.org/10.1002/ejsp.2111.

Viggers, H., Keall, M., Howden-Chapman, P., Wickens, K., Ingham, T., Davies, C., Chapman, R., Crane, J., 2019. Effect of an electricity voucher on electricity use. Energy Pol. 134, 110985 https://doi.org/10.1016/j.enpol.2019.110985.

Weigert, A., Hopf, K., Weinig, N., Staake, T., 2020. Detection of heat pumps from smart meter and open data. Energy Inform. 3 (21) https://doi.org/10.1186/s42162-020-00124-6.

Winkler, J., Das, S., Earle, L., Burkett, L., Robertson, J., Roberts, D., Booten, C., 2020. Impact of installation faults in air conditioners and heat pumps in single-family homes on U.S. energy usage. Appl. Energy 278, 115533. https://doi.org/10.1016/j.apenergy.2020.115533.

Wooldridge, J.M., 2010. In: Econometric Analysis of Cross Section and Panel Data, second ed. MIT Press, Cambridge, Mass. ed.

Yin, P., Pate, M., Battaglia, F., 2019. In-field performance evaluation and economic analysis of residential ground source heat pumps in heating operation. J. Build. Eng. 26, 100932 https://doi.org/10.1016/j.jobe.2019.100932.