CEP 19-10

Visualizing Energy Efficiency:
A Randomized Controlled Intervention

Maya Papineau
Carleton University

Nicholas Rivers
University of Ottawa

November 2019

CARLETON ECONOMIC PAPERS

Department of Economics

1125 Colonel By Drive
Ottawa, Ontario, Canada
K1S 5B6

Visualizing energy efficiency: A randomized controlled intervention*

Maya Papineau[†]
Carleton University

Nicholas Rivers[‡]
University of Ottawa

November 2, 2019

Abstract

We test the energy consumption impact of providing visual information on residential home heat loss with a social norm that informs households of their heat loss rate relative to their neighbours, and compare this to the impact of a traditional home energy report. Heat loss is visualized using infrared images taken from above approximately 14,000 households using a thermal image sensor mounted on a small aircraft during the winter heating season. Infrared images showing roof heat loss were provided to approximately 4,500 randomly selected households in on-bill messaging. A similarly-sized randomly selected group received bill messaging with a 'traditional' social norm comparing their consumption to similar homes. Both treatment groups were also shown a personalized estimate of the annual savings from reducing their consumption. Electricity and natural gas consumption are compared between treatment and control households during heating season over a one year period following the beginning of the intervention. After controlling for the estimated annual savings customers could achieve, natural gas consumption in the heat loss treatment falls by more than double the reduction in the traditional social norm, relative to control households. We conclude that home heat loss imaging and framing consumption in terms of heat loss hold promise in increasing the savings achieved from home energy reports.

Keywords: Energy efficiency, nudge

*We acknowledge the substantial contributions of Grant Guenther and Mike Dykstra at the City of Medicine Hat; Daygan Fowler and Darren Jones at MyHEAT; and funding from Natural Resources Canada and the Social Sciences and Humanities Research Council. Massimo Filippini, Ken Gillingham, and Arik Levinson as well as seminar and conference participants at Carleton University, McGill University and the Canadian Resource and Environmental Economics Association provided valuable comments.

[†]Department of Economics, Carleton University (maya.papineau@carleton.ca)

[‡]Graduate School of Public and International Affairs and Institute of the Environment, University of Ottawa (nicholas.rivers@uottawa.ca)

1 Introduction

Over the last decade, a number of energy distribution utilities worldwide have begun to use expanded information provision and behavioural nudges, such as Home Energy Reports (HERs), as a way to motivate consumers to reduce energy consumption (Allcott, 2011; Nolan et al., 2008; Schultz et al., 2007). HERs provide energy consumers with feedback that compares their own usage to that of comparable households, and also provide tips for saving energy. The most prominent example is OPOWER's (now Oracle's) Home Energy Reports (HERs), which at present are sent regularly to about 50 million customers at more than 100 utilities (Government of New Brunswick, 2017). These interventions are seen as a cost effective way to modestly reduce energy consumption with few barriers to implementation.

HERs and other similar "nudge" interventions are an effort to confront the "energy efficiency gap," which documents a persistent difference between the level of energy efficiency investments that appears to be profitable and those that actually occur. Popular explanations for the energy efficiency gap include market failures, such as incomplete information or credit constraints, and behavioural explanations, such as limited attention or the framing of choices (Gillingham and Palmer, 2014; DellaVigna, 2009). Home Energy Reports help to address these potential sources of the energy efficiency gap by making energy consumption more salient, increasing the information available to consumers, and framing energy consumption as a normative behaviour through comparisons of each customer's consumption to their neighbours' consumption. Recent work in this context has found messaging combined with pricing variation induces larger savings than moral appeals to conserve (Ito et al., 2018), price messaging combined with real-time consumption feedback increases the price elasticity of demand (Jessoe and Rapson, 2014), and pre-nudge heterogeneity in informedness about energy use partly explains why low consumption households tend to "boomerang" and increase consumption post information treatment (Byrne et al., 2018).

In this paper, we report the results of a randomized controlled trial that aims to test a behavioural nudge that builds on home energy reports. The intervention we test is novel, and is designed to elicit consumer response via both the informational and behavioural channels that have been identified as key explanations of the energy efficiency gap. The intervention involves providing consumers with high-resolution infrared images of their house on their monthly utility bills. The infrared images are taken at night from a small aircraft in the heating season in a Canadian city with a cold climate, and capture heat loss from customers' roofs. The thermal images clearly illustrate sources of heat loss in each house, and are accompanied with a personalized home heat loss score. If missing information is a contributor to the energy efficiency gap, this intervention could help to close the gap. Moreover, unlike a table or chart that accompanies typical Home Energy Reports (and standard energy bills), the thermal images convey energy consumption in a much more visual and salient manner, which could itself lead to a different response. In addition to the images, households are provided with information comparing their home to that of their neighbours. If consumers are "behaviourally" motivated by their energy consumption relative to their peers, this intervention helps to provide incentives to reduce energy consumption. Finally, like conventional HERs, recipients are provided with information about their potential bill savings from energy efficiency improvements as well as tips for improving energy performance.

A number of studies have found that traditional HERs are effective at reducing energy consumption. For example, Allcott (2011) finds that HERs rolled out by OPOWER starting in 2009 resulted in a reduction in electricity consumption of about 2 percent for an average household. Allcott and Rogers (2014) show that the electricity savings from these programs persist for several years, with substantial savings of energy persisting even after households stop receiving HERs. On the other hand, Allcott (2015) finds that utilities who were among the early adopters of OPOWER's home energy reports had larger shares of high income and environmentalist consumers, such that evaluations performed using data from early-adopter utilities, despite their high internally validity due to randomization, overstate future program efficacy in other customer populations. Allcott (2015) concludes that these results should focus researchers' efforts on recruiting utilities with more representative populations. Similarly, Costa and Kahn (2013) evaluate a randomized controlled trial of home energy reports, and find that politically liberal households respond to the intervention by reducing electricity consumption by 3.6%, while politically conservative households respond less than a third as much.

In addition to HERs, energy distribution companies and researchers have tested the impact of a number of other approaches to providing feedback to customers in order to inform and motivate reductions in energy consumption. For example, Ito et al. (2018) test the impact of moral suasion on household behaviour by exposing randomly selected households to messaging supportive of energy conservation. They find that there is an immediate impact following the treatment, but that it diminishes quickly after the intervention and that households do not continue to respond to repeated interventions. Martin and Rivers (2018) test the impact of providing in-home energy use displays to households, and find that households that receive such a device reduce electricity consumption by about 3 percent. Schleich et al. (2013) test the effect of providing web-based feedback to electricity consumers, and find a reduction in consumption of 4.5 percent.

While most HER deployments have focused on electricity, Allcott and Kessler (2019) study a HER program targeted at natural gas consumers. They find that the program induced a reduction in natural gas consumption of around 1 percent. Compared to the impact of HERs on electricity consumption, this natural gas savings figure is consistent with a broader finding that HERs induce smaller natural gas savings compared to electricity (Smith and Morris, 2014, Kerr and Tondro, 2012). This suggests that in cold countries with high heating demands that are met by fossil fuels, there are large potential gains from strategies that improve the effectiveness of HERs. For example in Canada, over 61 percent of end-use energy and 73 percent of residential sector greenhouse gases are from space heating (Natural Resources Canada, 2019), and close to 60 percent of space heating energy use derives from fossil fuels, mainly natural gas (Natural Resources Canada, 2018).[1]

Our study builds on this body of prior work and is targeted at improving the effect of HER social comparisons on heating usage. We assess whether visual heat loss information and messaging can improve upon the outcomes observed in the current HERs literature. Highly salient images of home energy loss may provide unique actionable information (for example, on an air leak) and induce some customers to respond in a way that numerical information on home energy consumption cannot replicate. In addition, a number of psychological studies provide a basis for the hypothesis that thermal images may provide behavioural cues that motivate consumers to take action to improve energy efficiency. In particular, thermal images may provide "vivid" representations of energy loss, which draw viewers in, hold attention, and excite the imagination, in a way that tabular information fails to do (Taylor and Thompson, 1982; Nisbett and Ross, 1980). Images are also considered to affect behaviour by being more available for recall during decision making, and to convert abstract ideas (energy loss) into more concrete terms that can be acted upon Sheppard et al. (2011). Two prior studies using small samples of voluntarily recruited participants in the UK offer suggestive evidence that improving the visibility of energy use may be more effective at reducing consumption compared to energy audits or textual information (Goodhew et al., 2015; Boomsma et al., 2016).

Our paper also contributes to a literature that documents divergences between ex-ante predicted savings from efficiency investments relative to realized, or ex-post, savings. A consensus is developing that, in some cases, engineering models' predictions about the returns to residential energy efficiency investments tend to overpredict returns, sometimes by a large margin (Levinson, 2016; Papineau, 2017; Fowlie et al., 2018). Our study utilizes what we term a hybrid engineering and realized savings approach to make predictions about energy consumption and bill reductions from improved building envelope energy efficiency. More specifically, our use of infrared technology to detect long wave thermal radiation mapped to individual homes allows us to identify residence-level heat loss, generate comparative metrics of the thermal attributes between different homes, then link these attributes to residential consumption data to predict annual bill savings from a reduction in heat loss (or equivalently, an improvement in building envelope energy efficiency). As discussed later in the paper, we find a strong relationship between measured heat loss rates and household natural gas consumption.

[1] These figures are calculated using secondary end-use energy consumption and therefore don't include transmission losses or the energy used to bring supplies to consumers.

For the study, we partnered with a municipally-owned natural gas and electricity distribution company. Roughly 14,000 single-detached households were randomly selected to participate in the study, and were randomly divided into two treatment groups and one control group of equal sizes. One treated group, which we call the 'traditional' HER group, were shown a figure comparing their electricity and natural gas consumption to both average and energy efficient similar homes, and their estimated annual savings from reducing their consumption. The second treatment group, which we call the thermal imaging group, received a high-resolution infrared image of their roof indicating heat loss, their thermal image-based heat loss score, how this score compares to that of their neighbours, and an estimate of annual savings from improving their heat score. Both treatments were also given the same set of four tips to reduce their consumption, which included both building envelope investments and behavioural changes.

We evaluate these interventions using daily data on natural gas and electricity consumption for all of the households in the control and both treatment groups. Our data covers the period from about one year prior to treatment to one year following initial treatment. Our analysis produces several findings. First, we find that for the typical customer, neither treatment produced a large impact on either electricity or natural gas consumption during the one-year period following the initiation of treatment. On average, the thermal imaging treatment caused consumers to reduce natural gas consumption by about 0.6%, but had small and statistically insignificant impacts on electricity consumption. Likewise, we found only small and statistically insignificant impacts on both natural gas and electricity consumption resulting from the traditional home energy report.

Our findings relating to the overall impact of these nudge-type interventions are fairly small compared to the extant literature. There are three potential explanations. First, Medicine Hat, where the experiment took place, is amongst the most conservative regions in Canada. In the 2019 federal election, right-leaning (Conservative and People's Party) parties garnered 82% of the vote in this electoral district, whereas the winning Liberal party only received a 6% vote share. Costa and Kahn (2013) show that (in the US) conservative (republican) voters respond three times less to HER interventions compared to progressive voters. Second, Medicine Hat is a key hub of Canada's natural gas industry (it is called "The Gas City"), and has one of the lowest shares of green voters in the country (2.3% in the 2019 federal election). Again, Costa and Kahn (2013) show that environmentalists respond much more to HER interventions compared to non-environmentalists. Third, Allcott (2015) shows that regions that adopt HERs early tend to have favourable conditions for their success, and have larger impacts than later-adopting regions. Medicine Hat is a "later-adopting region," and may be expected to have lower impacts from such an intervention that early-adopting regions.

However, further analysis exposes substantial heterogeneity within each of the treatment groups: while there is only a small aggregate impact, we find that low-efficiency/high consumption households respond by reducing electricity and natural gas consumption by more than 5 percent on average, while high efficiency/low consumption households respond by increasing usage by roughly 3 percent, a "boomerang" effect that has been documented in other studies (Byrne et al., 2018; Delmas et al., 2013).

For both treatments, households were informed of how much they could save annually on energy bills by improving their energy efficiency to a given level. In the traditional social comparison treatment group, the estimated dollar savings were based on a comparison between household energy consumption and mean consumption of comparison (similar) households. Households that consumed more than average were told that they could save money (and told how much) by improving to the average level, whereas households that consumed less than average were told that they were savings money (and told how much) as a result of having lower than average energy consumption. In the thermal image treatment group, the dollar savings were estimated as the savings that would result from improving the house's thermal imaging score to the best possible category. When we take into account the heterogeneity in potential savings, we find that while both treatments had a statistically significant effect on natural gas consumption per dollar of estimated savings, the heat loss social comparison had a larger effect. At the mean annual heat loss group estimated savings of $150, the traditional social comparison reduced natural gas consumption by 1.8 percent, whereas the heat loss treatment reduced consumption by 3.9 percent - more than double the traditional HER.

Finally, by linking the household addresses to a database maintained by the provincial energy efficiency agency, we show that households that receive the heat loss social comparison treatment are more likely to participate in energy efficiency programs following treatment than either the control group or the households in the traditional home energy report group. These energy efficiency programs are targeted at improving the thermal integrity of the building shell and improving the efficiency of household appliances. As a result, these results are suggestive that the intervention produced gains in energy efficiency and not just transient changes in behaviour.

The rest of the paper is organized as follows. Section 2 begins with a brief overview of recent contributions on behavioral "nudges" in the context of energy conservation, then describes our experimental treatments that are the focus of this paper. Section 3 provides an assessment of the relationship between our measured heat loss ratings and realized energy consumption. Section 4 describes our data sources for the variables used in the analysis, and Section 5 presents our results. Section 6 briefly concludes.

2 Overview of context and experiment

The experiment takes place in Medicine Hat, Alberta. Medicine Hat is a city of about 60,000 located in southeast Alberta, with relatively hot summers and cold winters. The municipally-owned utility provides gas, electricity, and water to residents and businesses, and was responsible for coordinating the experiment whose results we report here.[2] Until this experiment, the City had not implemented other behavioural feedbacks on energy bills, such as Home Energy Reports.

The experiment consists of providing on-bill feedback to randomly selected households in Medicine Hat. Households in Medicine Hat receive monthly utility bills (including natural gas, electricity, and water), and the intervention began by including the treatments on the February 2018 billing cycle.[3] The intervention was repeated on the March, April and November billing cycles of 2018. Figure 1 illustrates the timeline of the experimental intervention.

The intervention was run by the City of Medicine Hat as a *natural field experiment*. As described in Czibor et al. (2019), in a natural field experiment subjects include the relevant population (rather than a sub-sample of voluntary participants) and subjects are not aware of being part of an experiment, which eliminates Hawthorne effects that have been shown to be important in similar studies (Schwartz et al., 2013).

The experiment population includes all municipally-served single-family residential buildings in Medicine Hat. This population was randomly assigned to two experimental groups and a control group using a max-min t-statistic re-randomization algorithm to ensure balance (Bruhn and McKenzie, 2009).[4] Table 1 shows balance statistics for the three groups, and confirms that the randomization delivers groups that are balanced on observable covariates. For each experimental group, the t-statistic comparing the treatment group with the control group is well below the critical value of 1.96.

Appendix B shows calculations of statistical power that were conducted in support of the development of the experiment. Based on prior analyses of energy feedback interventions, the statistical power calculations show that our experimental design has a high power to recover the effects of treatment if they are of similar magnitude to estimates from prior interventions.

The first treatment (hereafter referred to as treatment 1) is similar to the Home Energy Reports that have been used extensively in recent years to inform households about their energy consumption relative to that of their neighbours (Allcott, 2011; Ayres et al., 2013). Households in this 'traditional' HER treatment group received on-bill messaging that includes a month-to-

[2] Although in principle the municipality is able to set its own gas and electricity rates, in practice rates are set at the average of the rates for other provincially-regulated utilities.

[3] Like most utilities, Medicine Hat stratifies its customers into groups, who are on different billing cycles, so not all treated households receive the treatment on the same day.

[4] The balancing variables include pre-intervention gas and electricity consumption, year of construction, assessed value, and building size.

month consumption comparison between a given household and the 50 most similar households, as well as the top quintile of most efficient households among the group of similar households.[5] Large text in a yellow box on the first page of the bill provides a comparison between the household energy consumption in the prior month and energy consumption in the group of similar homes in the same month, with an indication of potential bill savings from reducing energy consumption to the level as that of the mean of similar households. For households with consumption above the mean consumption of similar households, this number is presented as a potential savings from improvement in energy efficiency, whereas for households with consumption below the mean consumption of similar households, this value is presented as the savings achieved from having a high level of energy efficiency. This page also prompts customers to see more detail on their relative consumption on page 4 of their bill, with the statement 'See page 4 for your personalized comparison and options to save energy'. Page 4 of the bill presents graphical information on natural gas and electricity consumption over the past year for the household compared to similar households. A list of potential options for reducing energy consumption are also included on page 4. A sample bill for this treatment is included in Appendix C.1.

Households in treatment group 2 were provided with infrared images of their roof. The infrared images were taken at night in the heating season three months before the experiment, and measure heat loss from the home's roof. Thermal images were acquired using the MyHEAT technology platform, which is a combination of image acquisition equipment and processing software designed for the purpose of measuring heat loss from buildings.[6] Thermal images are gathered using an aircraft-mounted thermal infrared sensor, which detects emitted long-wave radiation. These images are used in conjunction with other measurements (e.g., temperature, elevation) as well as building shapefiles to create a thermal profile for all buildings in a municipality. The combined process is able to produce extremely high-resolution thermal images of building roof heat loss, accurate to within 0.05°C at a sub-one metre resolution. Using the thermal images, each dwelling is assigned a heat loss score ranging from 1 to 10, which indicates the amount of heat loss from the roof: 1 indicates very low heat loss; 10 indicates very high heat loss.[7] Thermal images corresponding to each of the possible heat loss scores are provided in Figure 2. The following section of the paper uses pre-program energy consumption data to verify that the thermal images and associated image-based heat loss ratings convey meaningful

[5] 'Similar' homes were the group of 50 homes with the smallest differences with the comparison home in terms of year built and size. 'Similar' homes were also restricted to homes that were on the same billing cycle as the treated household. More precisely, define an index of similarity (IS) between a target house j and a possible comparison house h, which compares the attributes of house h to attributes of the target house j: $IS_{hj} = (Nsize_h - Nsize_j)^2 + (Nyear_h - Nyear_j)^2$, where $Nsize_h = (size_h - mean(size))/sd(size)$ and $Nyear_h = (year_h - mean(year))/sd(year)$. The group of similar homes consists of the 50 homes with the smallest index of similarity.

[6] See https://MyHEAT.ca/technology.

[7] The heat loss scores are assigned using a proprietary algorithm developed by MyHEAT.

information about the relative energy performance of dwellings in the experimental region.

Households in treatment group two were also provided with on-bill messaging, which include text on the front page informing them of their heat loss performance and potential bill savings from improving their MyHEAT score. Potential bill savings were calculated using the regression coefficients estimated in the following section, and were determined based on the energy savings from an improvement to a heat loss score of 1 (the best possible score). This text is accompanied by a prompt to find further information on page four, as in treatment one. However, the page four information differs from the standard social comparison in treatment group one. It includes the thermal image of their house, along with brief instructions for interpreting the image. Households were also provided with their heat loss score, along with the average heat loss score for houses in their neighbourhood and the average heat loss score for houses in the City of Medicine Hat. Finally, households were provided with the same list of potential options for reducing home energy consumption as in treatment group one. A sample bill is included in Appendix C.2.

Households in both treatment groups received on-bill messaging for the first time starting in February 2018. Households were provided with messaging for three consecutive months in February through April 2018. Another bill treatment was included in the November 2018 bill.[8] These months were chosen as they cover the heating season, when building heat loss is most important for determining energy consumption.

The treated unit in the analysis is the physical location, and so we do not have concerns about attrition from the experiment.[9] In addition, we construct a balanced sample by only including locations for which consumption data is available over the entire analysis period to ensure that entry has no effect on our results.

In the following section, we compare pre-intervention natural gas and electricity consumption across homes with different thermal images. We show that buildings with higher (worse) heat loss ratings consume substantially more natural gas, as well as somewhat more electricity, compared to similar homes with lower (better) heat loss ratings.

3 Relationship between MyHEAT rating and energy consumption

This section documents the relationship between infrared image-based MyHEAT ratings and building energy consumption. To do this, MyHEAT scores, which were collected on October 31 2017, are compared to building energy consumption data for 2015 and 2016. For each dwelling in the data, we merge monthly billed energy consumption in 2015 and 2016, MyHEAT rating,

[8] The February mailout was provided in color, while subsequent mailouts were black and white.

[9] We have data on customers, and this indicates that a minimal number of customers move during the analysis period.

and tax assessment information. Tax assessment information includes data on building size, building type,[10] year built, assessed value, as well as neighborhood and street name. The full merged data set consists of 14,373 observations. Each observation represents a single dwelling, and contains the MyHEAT score, average annual natural gas and electricity consumption over 2015 and 2016, and building characteristics.[11]

Table 2 summarizes the data. HEATSCORE and NEIGHSCORE are the dwelling and neighbourhood MyHEAT scores. Electricity is the average annual electricity consumption in kWh over 2015 and 2016. Gas is the average annual gas consumption in GJ over 2015 and 2016. Building Size is the assessed size of the building in m^2. Assessed Value is the assessed building value and Year Built is the year the building was constructed. In addition to the numeric variables recorded in Table 2, we also observe the neighbourhood name, street name, and building type.

We assess the relationship between MyHEAT rating and building energy consumption using a regression framework. We consider a cross-sectional regression of the form:

$$\log(Y_i) = \beta_0 + \beta_1 \text{HEATSCORE}_i + \beta_2 \log(\text{buildingSize}_i) + \beta_3 X_i + \epsilon_i, \tag{1}$$

where i indexes dwellings, HEATSCORE_i is the MyHEAT score, Y_i is average annual energy (electricity or gas) consumption, buildingSize_i is assessed building size, and X_i includes other observable variables, such as building age, building type, etc.

Table 3 summarizes the results of the analysis in which natural gas consumption is the dependent variable. Column (1) includes a control for building size only. The estimate suggests that a one unit improvement in the MyHEAT score, or equivalently a reduction in measured heat loss, is associated with a 4.1 percent reduction in natural gas consumption. The second column also includes building type as an explanatory variable, such that only buildings of the same type are compared to each other. This column suggests that a one unit reduction in the MyHEAT score is associated with a 4.7 percent reduction in natural gas consumption. The third column adds controls for year built. This column suggests that each one-unit MyHEAT score improvement reduces gas consumption by 3.1 percent. The fourth column adds a control for neighbourhood name and street name. The fifth column adds a control for the (log of the) assessed value of the house. In each of these last three columns, the coefficient remains unchanged. Based on these estimates, we estimate that each one unit improvement in the MyHEAT score is associated with

[10] There are 16 building types in the assessment data with at least 100 observations: 1 1/2 storey with basement, 1 3/4 storey with basement, 1 storey duplex with basement, 1 storey multi side x side basementless, 2 storey basementless, 2 storey duplex with a basement, 2 storey multi side x side with basement, 2 storey with basement, bi-level, bungalow basementless, bungalow with basement, duplex bi-level, mobile home double wide foundationless, mobile home single wide foundationless, no market building class, split level.

[11] We remove 21 dwellings for which the building size is listed as less than 10m^2, as well as 3,086 observations (or roughly 1,500 dwellings) for which we are missing a MyHEAT score or for which annual gas, electricity, or water consumption is zero over either 2015 or 2016.

3.1-4.7 percent less natural gas consumption. In each case, the standard errors indicate we are able to estimate the effect with a substantial amount of precision and reject the null hypothesis that there is no relationship between MyHEAT rating and energy consumption.

Table 4 estimates the same regression using electricity consumption, rather than gas, as a dependent variable. Aside from the first column, which does not include controls for anything except building size, the regression coefficients suggest each unit of MyHEAT improvement is associated with savings of electricity of 1.3-1.8 percent. Once again, the effect is estimated precisely.

Tables 3 and 4 treat the MyHEAT rating as a continuous variable, and find that reductions in the MyHEAT score of a dwelling (i.e., decreases in measured heat loss) are associated with reductions in energy consumption. In Figure 3, we re-estimate the models above, but treat the MyHEAT rating as a discrete variable:

$$\log(Y_i) = \beta_0 + \sum_{n=1}^{10} \beta_n \mathbb{1}(HEATSCORE_i == n) + \beta_2 \log(SIZE_i) + \theta X_i + \epsilon_i \qquad (2)$$

where $n = 1..10$ indicates the set of possible MyHEAT ratings. We treat dwellings with a MyHEAT rating of 5 as the reference category, and measure energy consumption relative to that category. We adopt the formulation in column (2) of the tables above, which conditions energy consumption on both building size as well as building type.

Consistent with the prior analysis, the figures show a strong relationship between natural gas consumption and MyHEAT rating. Buildings with a MyHEAT rating of 10 consume on average about 50% more natural gas than similarly sized buildings of the same type with a MyHEAT rating of 1. For electricity, the results show a distinct relationship between MyHEAT rating and electricity consumption as well, although the standard errors are larger, particularly on houses with extreme MyHEAT ratings, such that the relationship is not as clear as for gas. This is not surprising, since the primary space heating fuel is natural gas, rather than electricity, in the city under study.

Overall, the findings in this section show that the MyHEAT rating is a useful predictor of residential energy consumption.

4 Experiment Data

We combine data from a number of sources to conduct the analysis that follows. Altogether, after cleaning data, we observe bills and consumption from 13,870 households. The control group includes 4,565 households, treatment group one includes 4,642 households, and treatment group two includes 4,663 households.

Our dataset is constructed on the basis of a number of different data sources. The first data source is monthly consumption and expenditure data for each household and billing period. Utility bills provide monthly information on natural gas, electricity, and water consumption, and are available starting in 2015.

Second, in addition to monthly billing data, we also obtain a separate source of consumption data from household meters. Medicine Hat uses digital (smart) electricity meters that record electricity and natural gas consumption at both daily and hourly intervals. Our main analysis is based on daily natural gas and electricity consumption data. The consumption data starts in January 2017 and extends until March 2019, roughly one year prior to and after treatment (see Figure 1). We retain only households with a complete set of daily consumption data, such that there is no entry to the sample nor attrition from the sample.

Third, we obtain tax assessment data, which provides information on building size, assessed value, building type (e.g., split level, bungalow, etc.), neighbourhood, and year of construction.

Fourth, we obtain thermal imaging data for all residential dwellings from MyHEAT. As described above, each dwelling is given a heat loss score ranging from 1 (low heat loss; high efficiency) to 10 (high heat loss; low efficiency). We observe this score for all buildings in the population.

The unit of observation in our analysis is the residential building-day, rather than the customer. This is useful since it makes it straightforward to ensure that there is no attrition from or selection into our sample. However, it does not take into account possible moving of customers into and out of houses during the period covered by our analysis. To the extent that households respond to the intervention with behavioral or habit changes, rather that permanent physical changes to the housing equipment or envelope, moves of customers between households after the treatment is initiated will attenuate the treatment effect we seek to estimate.

5 Analysis

5.1 Impacts on natural gas and electricity consumption

We begin the analysis by using the data described above to estimate a model that captures the impact of treatment on energy consumption:

$$Y_{it} = \beta_0 + \sum_{k=1}^{2} \beta_k T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}, \tag{3}$$

where Y_{it} is consumption (either gas or electricity), normalized by average post-treatment consumption in the control group.[12] T_{ik} is a treatment dummy, which indicates whether

[12]This normalization is the same as that used by Allcott (2011). We use normalization instead of taking logs to

household i is in treatment group $k = \{1,2\}$, and P_{it} is a post-treatment dummy that indicates whether the observation is in the post-treatment period (note that since households are mailed bills on different dates, the post-treatment period differs by household). We define the post-treatment period as any time after the mailout of the first treated bill to the household. We also include a location (house) fixed effect μ_i and day-of-sample fixed effect λ_t. The location fixed effect absorbs any persistent differences between households (number of occupants, thermal properties of the dwelling, etc.). The day-of-sample fixed effect absorbs common factors that shift over time that impact households (weather, holidays, etc.). In specification (3) β_k is then the average effect of treatment k – the effect of treatment on electricity or natural gas consumption in the post-treatment period. Given our specification in equation (3), β_k is identified from within-household and within-day differences between the treatment and control groups. Throughout the paper, standard errors are two-way clustered by household and day-of-sample.

We report the results of estimating (3) in Table 5. Columns (1) and (2) pertain to the impact of the interventions on daily natural gas consumption, and columns (3) and (4) describe the impact on daily electricity demand. Columns (1) and (3) aggregate the impacts of the traditional HER and and the thermal image interventions together, and columns (2) and (4) separately identify the effects of each of these interventions, relative to the control group. The results in columns (1) and (3) indicate that the two treatments combined had a very small to no impact on either natural gas or electricity consumption on average. The coefficients in column (2) show that on average the traditional HER treatment did not significantly reduce daily natural gas consumption. In contrast, the coefficient on the MyHEAT thermal imaging intervention is almost an order of magnitude larger and significant at the 90 percent confidence level. In the case of electricity, column (4) indicates that both treatment effects are small and neither are statistically different from zero. In total, it is fair to say that the average effects of both treatments are quite small, and on the low side of other estimates in the literature reviewed in the introduction.

Specification (3) groups all households in treatment group 1 together, and all households in treatment group 2 together. However, there are important within-group differences in treatments that these households receive. In particular, part of both treatments is an estimate of the potential monetary savings from improving household energy efficiency and are provided with social cues relating to how much energy they consume relative to their neighbours. In treatment group 1, households are told how much they would save (or are already saving) on an annual basis if their energy consumption was the same as their average neighbour. In treatment group 2, households are told how much they could save on an annual basis if they were able to

avoid dropping zero consumption observations. Coefficients can be interpreted identically to a model with a logged left-hand side variable. In practice, we find little difference between the results of a model estimated with a normalized left hand side and a model estimated with a logged left hand side variable.

improve their MyHEAT rating to 1 (the best possible rating). In both treatments, households are also provided with normative cues that rank their energy consumption relative to that of their neighbours. Clearly, households that are informed that there are large potential savings from improvements in energy efficiency and that they are large energy consumers relative to their neighbours may respond differently to treatment than households who are told there are small (or negative) savings and that they consume less than their neighbours.

We evaluate the hypothesis that treatment effects are heterogeneous depending on messaging received by estimating the following equation,

$$Y_{it} = \sum_{g=1}^{3} T_{i1} P_t \theta_i^g + \sum_{h=1}^{3} T_{i2} P_t \theta_i^h + \mu_i + \lambda_t + \epsilon_{it}. \tag{4}$$

where θ_i^g is a set of dummy variables that allocate each household in treatment 1 into one of three groups g corresponding to whether households saw messaging that they were spending more, less or the same on their annual utility bills relative to the average of similar households, and θ_i^h is a set of dummy variables that groups households in treatment 2 into low, medium, and high, based on their MyHEAT scores.[13] We report the results from estimating this equation in Figures 4 and 5.

In Figure 4, 'Negative' denotes customers who were told they were consuming less than average and therefore saving on billing expenditures relative to similar households; 'Zero' denotes customers who were told they were saving zero dollars relative to similar households; and 'Positive' denotes customers who were consuming more than similar households and therefore told they were paying more money than average. For both gas and electricity consumption, we observe that customers who were told they were saving money relative to the average household increased their consumption, whereas customers who were told they were spending more money relative to the average household increased their consumption.

In Figure 5 households with the highest MyHEAT scores respond to treatment by reducing natural gas consumption, whereas households with the lowest MyHEAT scores respond to treatment by increasing gas consumption. A similar pattern is observed for electricity consumption, though the relative changes are not statistically significant.

Both of these figures confirm our hypothesis that the treatment effect is heterogeneous within each program, depending on specific messaging received. This is a finding that has been reported in the literature previously, and is sometimes referred to as the "boomerang effect," in which more efficient households actually increase their consumption in response to comparisons with their neighbours that make clear their relative efficiency (Schultz et al., 2007).

[13] The low group includes MyHEAT ratings 1-3; the medium group includes MyHEAT ratings 4-7; and the high group includes MyHEAT ratings 8-10.

Figures 4 and 5 are useful for illustrating the heterogeneity in responses to different messaging in each intervention. However, they don't allow for the two treatments to be directly compared against one another. To do this, we estimate a model in which we interact each treatment with the dollar value of savings each household was told they could save on energy if they made improvements in their dwelling. In particular, specification (5) controls for the heterogeneity in each household's estimated annual savings by including variable D_{ikm}, the dollar savings estimate household i in treatment group k were shown on both pages one and four of their utlility bill in billing month m, in units of hundreds of dollars. The interpretation of coefficient α_k is the percent reduction in consumption in treatment k per hundred dollars of estimated savings.

$$Y_{it} = \alpha_0 + \sum_{k=1}^{2} \alpha_k D_{ikm} \times T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}. \tag{5}$$

Specification (6) allows for a heterogeneous response to treatment by estimated dollar savings as above, but also allows for a potential response unrelated to the dollar savings, by incorporating the term $T_{ik} \times P_{it}$. The interpretation of coefficient ψ_k is the percent reduction in consumption in treatment k when dollar savings are zero.

$$Y_{it} = \alpha_0 + \sum_{k=1}^{2} \alpha_k D_{ikm} \times T_{ik} \times P_{it} + \sum_{k=1}^{2} \psi_k \times T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}. \tag{6}$$

We report the results from estimating equations (5) and (6) in Table 6. Columns (1) and (2) estimate the impact on natural gas consumption, and columns (3) and (4) on electricity. Columns (1) and (3) include only the interaction term between dollars and treatment (equation (5)), and columns (2) and (4) include both an interaction term and a main effect of treatment, as in (6).

In column (1) of Table 6, we report that the traditional HER treatment reduced daily natural gas consumption by 1.2 percent per hundred dollars of estimated savings, and the MyHEAT treatment reduced daily natural gas consumption by about 2.6 percent per hundred dollars of estimated savings. Both of these estimates are significant at the 99 percent level of confidence and the MyHEAT effect is statistically significantly larger than the traditional HER.[14] To put this in context, at the mean estimated MyHEAT annual potential saving of \$150, the traditional HER reduced gas consumption by 1.8 percent, whereas the MyHEAT treatment reduced gas consumption by 3.9 percent, more than double the traditional social comparison. For the

[14]The p-value for the test of the hypothesis that the traditional HER and MyHEAT estimates are equal is less than 0.1 percent.

electricity results in column (3) of Table 6, the MyHEAT social comparison brought about larger, statistically significant reductions in daily electricity consumption per hundred dollars of savings, whereas the traditional HER treatment led to very small, statistically insignificant reductions.[15]

A similar general pattern is observed in columns (2) and (4) of Table 6, where we estimate specification (6). In column (2), the traditional HER treatment reduced daily natural gas consumption by about 1.2 percent per hundred dollars of estimated savings, and the MyHEAT treatment reduced daily natural gas consumption by 7.5 percent per hundred dollars of estimated savings. The traditional social comparison for customers who were shown savings of zero reduced gas consumption did not change their consumption, however MyHEAT customers who were shown savings of zero (i.e., the most efficient households) exhibited a substantial boomerang effect by increasing their consumption by 10.5 percent. In column (4) of Table **??**, the MyHEAT treatment brought about daily electricity reductions of 2.7 percent per hundred dollars of positive estimated savings, whereas households who were shown savings of zero rebounded by increasing consumption by 4 percent per hundred dollars of savings. The traditional social comparison did not significantly affect electricity consumption.

The column (2) and (4) result discussed above suggest that HER treatments with heat loss messaging and imagery targeting relatively inefficient households hold promise in increasing gas savings relative to traditional HERs. The same is true for electricity, though to a more muted extent. The heterogeneity we observe among households who are informed that they are relatively energy efficient, implying they have less potential to gain from reducing consumption, respond differently from households who are less efficient, is not surprising. This type of heterogeneity in the impact of information on energy consumption has been documented in the prior literature (Byrne et al., 2018; Allcott, 2011; Costa and Kahn, 2013).

5.2 Impacts on energy efficiency program participation

In addition to examining impacts of treatment on gas and electricity consumption, we extend our analysis by estimating whether treatment caused households to participate in other energy efficiency programs. The on-bill treatments (both the traditional social comparison as well as the MyHEAT treatment) were accompanied with information about provincial energy efficiency programs that were available to the household. These energy efficiency programs are aimed at improving the energy efficiency of the household, principally by providing investment subsidies for high-efficiency equipment, insulation, and home energy audits. As a result, participation in these programs should be expected to induce an on-going reduction in energy consumption. To our knowledge, most evaluations of Home Energy Reports focus on immediate changes

[15]For both gas and electricity, an F-test reveals that the difference between the HER and MyHEAT treatment is statistically significant at the 99 percent confidence level.

in energy consumption, and have not evaluated impacts on durable good purchase or energy efficiency program participation, so these results are an important contribution to the literature.

We evaluate the impact on energy efficiency program participation using data on program participation rates from Energy Efficiency Alberta. The provincial agency provided us with program participation information for all households in Medicine Hat, linked to the treatment groups by matching on addresses.[16] Program participation information is for 2018, after the treatment is initiated for most households. We analyze the results using a regression-based approach, in which we regress a program participation dummy on a treatment indicator. Results are provided in Table 7.

The first column evaluates the impact of treatment on all energy efficiency programs offered by Energy Efficiency Alberta. As indicated in the table, participation in the control group is 6.1%. Participation in Treatment group 1 is 0.7% lower, but this difference is not statistically significant. Participation in the MyHEAT group is 1.0% percentage points higher, or about 17% higher (the difference is significant at the 10% level). Thus, there is evidence that the MyHEAT program causes increases in uptake of energy efficiency programs.

Energy Efficiency Alberta divides their residential-focused programs into three streams: (i) home improvement, which includes insulation and window improvement, (ii) online rebates, which include rebates for clothes washers, smart thermostats, and other household equipment, and (iii) home energy plans, which include funding for home energy audits. Given the focus of the MyHEAT program, we expect most of the increase uptake of energy efficiency programs to be in the Home improvement stream. Table 7 confirms that this is what we find. The second column shows participation in the home improvement stream, which is predominantly funding for insulation improvements. In the control group, 2.9% of households participate in this program. Participation rates in Treatment group 1 (the home energy report) are similar. Particiaption rates in the MyHEAT group are 0.7 percentage points higher, or about 25% higher. This large difference suggests that the MyHEAT treatment induced households to undertake insulation upgrades. The final two columns of Table 7 show participation rates in the other program streams. We do not find any evidence of increases in participation rates in the other streams (the Home Energy Report group actually participated less in online rebate programs relative to the control group, a finding that we cannot explain).

6 Conclusion

This paper reports on a randomized controlled trial that compares the effects of two different social comparisons among customers of a natural gas and electric utility in Canada: traditional

[16] An 85% success rate in matching was achieved, such that 15% of energy efficiency program participants could not be matched to a treatment group.

home energy reports versus visual imagery and messaging on home heat loss. Both treatments also included personalized messaging on the estimated annual billing expenditures they could save (or were saving) relative to energy efficient homes, from reducing their consumption.

We find that the heat loss treatment led to significantly larger consumption reductions per dollar of estimated savings, relative to the traditional home energy report social comparison. This is strongly suggestive evidence that home heat loss imaging along with framing consumption in terms of heat loss hold promise in increasing the savings achieved from the home energy reports that have become ubiquitous among utility customers in North America and Europe.

References

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

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

_ **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 2019, *11* (1), 236–276.

_ **and Todd Rogers**, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 2014, *104* (10), 3003–37.

Ayres, Ian, Sophie Raseman, and Alice Shih, "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage," *The Journal of Law, Economics, and Organization*, 2013, *29* (5), 992–1022.

Boomsma, Christine, Julie Goodhew, Steve Goodhew, and Sabine Pahl, "Improving the visibility of energy use in home heating in England: Thermal images and the role of visual tailoring," *Energy Research & Social Science*, 2016, *14*, 111–121.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Byrne, David P., Andrea La Nauze, and Leslie A. Martin, "Tell me something I don't already know: Informedness and the impact of information programs," *Review of Economics and Statistics*, 2018, *100* (3), 510–517.

Costa, Dora L and Matthew E Kahn, "Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment," *Journal of the European Economic Association*, 2013, *11* (3), 680–702.

Czibor, Eszter, David Jimenez-Gomez, and John A List, "The Dozen Things Experimental Economists Should Do (More of)," Technical Report, National Bureau of Economic Research 2019.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–372.

Delmas, Magali A., Miriam Fischlein, and Omar I. Asensio, "Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012," *Energy Policy*, 2013, *61*, 729–739.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram, "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program," *The Quarterly Journal of Economics*, 2018, *133* (3), 1597–1644.

Gillingham, Kenneth and Karen Palmer, "Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence," *Review of Environmental Economics and Policy*, 2014, *8* (1), 18–38.

Gleerup, Maria, Anders Larsen, Søren Leth-Petersen, and Mikael Togeby, "The effect of feedback by text message (SMS) and email on household electricity consumption: experimental evidence," *The Energy Journal*, 2010, pp. 113–132.

Goodhew, Julie, Sabine Pahl, Tim Auburn, and Steve Goodhew, "Making Heat Visible: Promoting Energy Conservation Behaviors Through Thermal Imaging," *Environment and Behavior*, 2015, *47* (10), 1059–1088.

Government of New Brunswick, "NB Power launches home energy reports for residential customers," January 2017.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka, "Moral suasion and economic incentives: Field experimental evidence from energy demand," *American Economic Journal: Economic Policy*, 2018, *10* (1), 240–67.

Jessoe, Katrina and David Rapson, "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, 2014, *104* (4), 1417–38.

Kerr, R. and M. Tondro, "Residential feedback devices and programs: Opportunities for natural gas," 2012. Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy.

Levinson, Arik, "How Much Energy Do Building Energy Codes Save? Evidence from California Houses," *American Economic Review*, 2016, *106* (10), 2867–94.

Martin, Steve and Nicholas Rivers, "Information provision, market incentives, and household electricity consumption: evidence from a large-scale field deployment," *Journal of the Association of Environmental and Resource Economists*, 2018, *5* (1), 207–231.

Natural Resources Canada, "Energy Fact Book," 2018.

_, "National Energy Use Database, Natural Resources Canada," 2019.

Nisbett, Richard E and Lee Ross, "Human inference: Strategies and shortcomings of social judgment," 1980.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "Normative Social Influence is Underdetected," *Personality and Social Psychology Bulletin*, 2008, *34* (7), 913–923. PMID: 18550863.

Papineau, Maya, "Setting the Standard? Evaluating the Cost-Effectiveness of Building Energy Standards," *Energy Economics*, 2017, *64*, 63–76.

Schleich, Joachim, Marian Klobasa, Sebastian Gölz, and Marc Brunner, "Effects of feedback on residential electricity demand—Findings from a field trial in Austria," *Energy Policy*, 2013, *61*, 1097–1106.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 2007, *18* (5), 429–434. PMID: 17576283.

Schwartz, Daniel, Baruch Fischhoff, Tamar Krishnamurti, and Fallaw Sowell, "The Hawthorne effect and energy awareness," *Proceedings of the National Academy of Sciences*, 2013, *110* (38), 15242–15246.

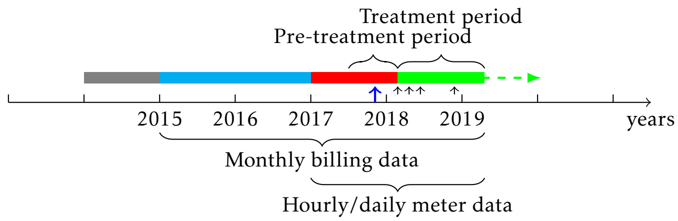
Sheppard, Stephen RJ, Alison Shaw, David Flanders, Sarah Burch, Arnim Wiek, Jeff Carmichael, John Robinson, and Stewart Cohen, "Future visioning of local climate change: a framework for community engagement and planning with scenarios and visualisation," *Futures*, 2011, *43* (4), 400–412.

Smith, Brian Arthur and Lucy Morris, "Neighbor comparison reports produce savings, but HOW?," 2014. ACEEE Summer Study on Energy Efficiency in Buildings.

Taylor, Shelley E and Suzanne C Thompson, "Stalking the elusive" vividness" effect.," *Psychological review*, 1982, *89* (2), 155.

A Tables and Figures

Figure 1: Timeline of experiment.



Notes: The small black arrows indicate when bill inserts were included in monthly household energy bills. The larger blue arrow indicates when the thermal images were gathered.

Figure 2: Examples of thermal images for buildings with heat scores from 1 (low energy consumption) to 10 (high energy consumption).

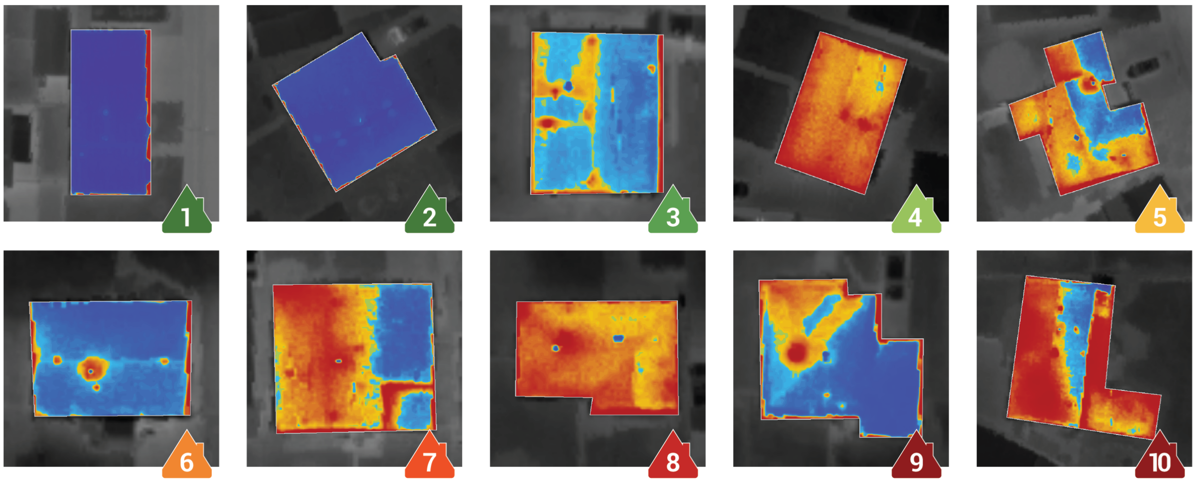
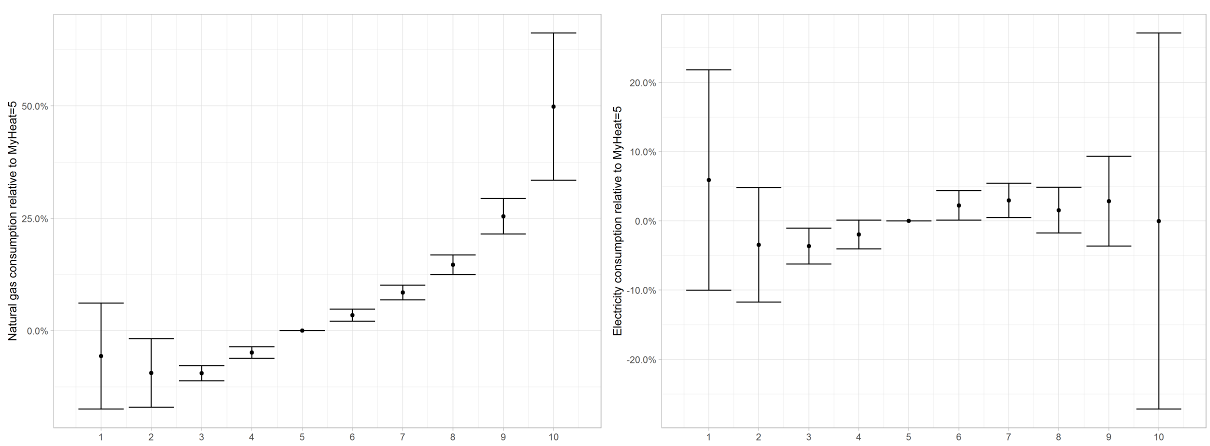
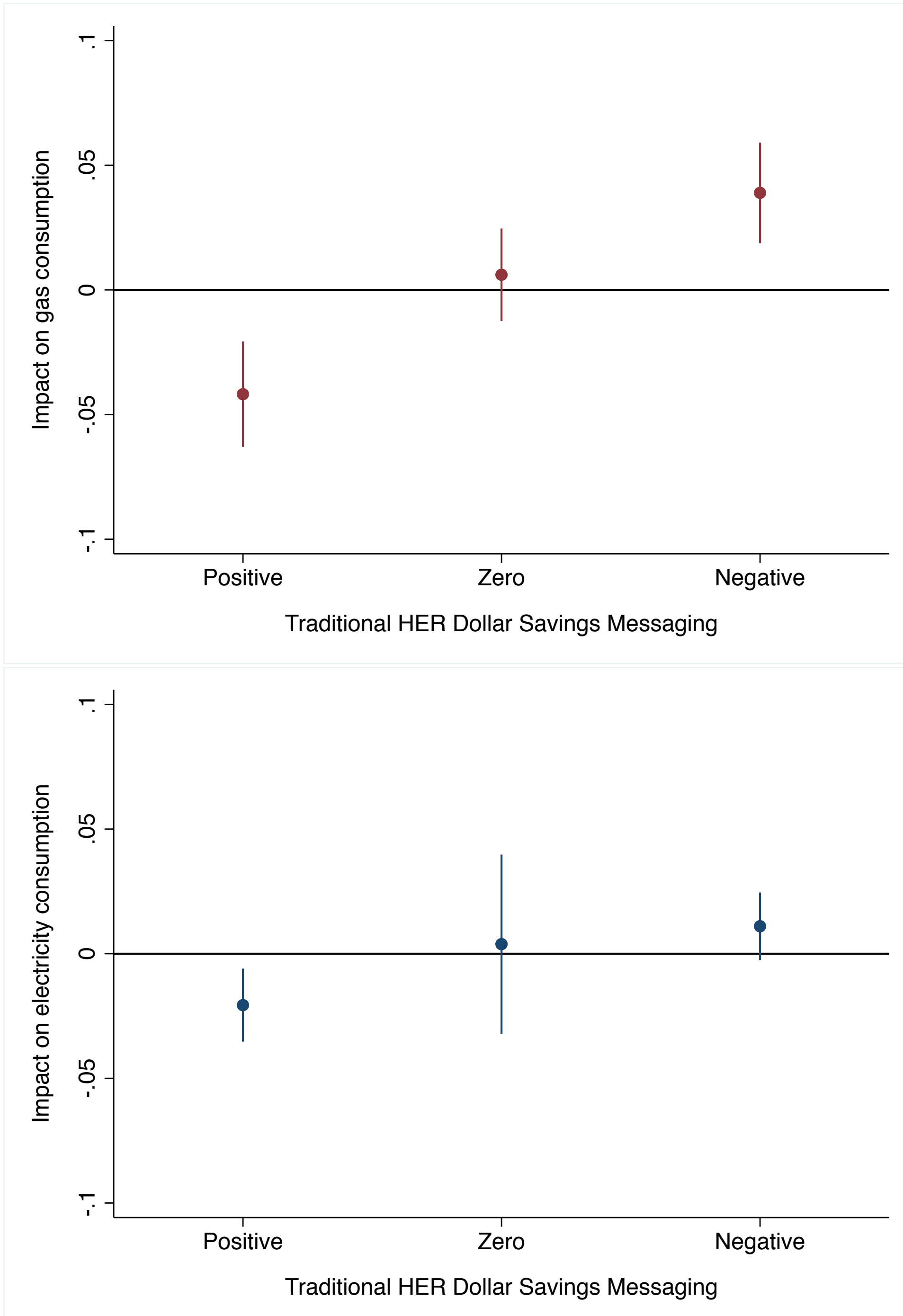


Figure 3: Comparison of MyHEAT rating and building energy consumption



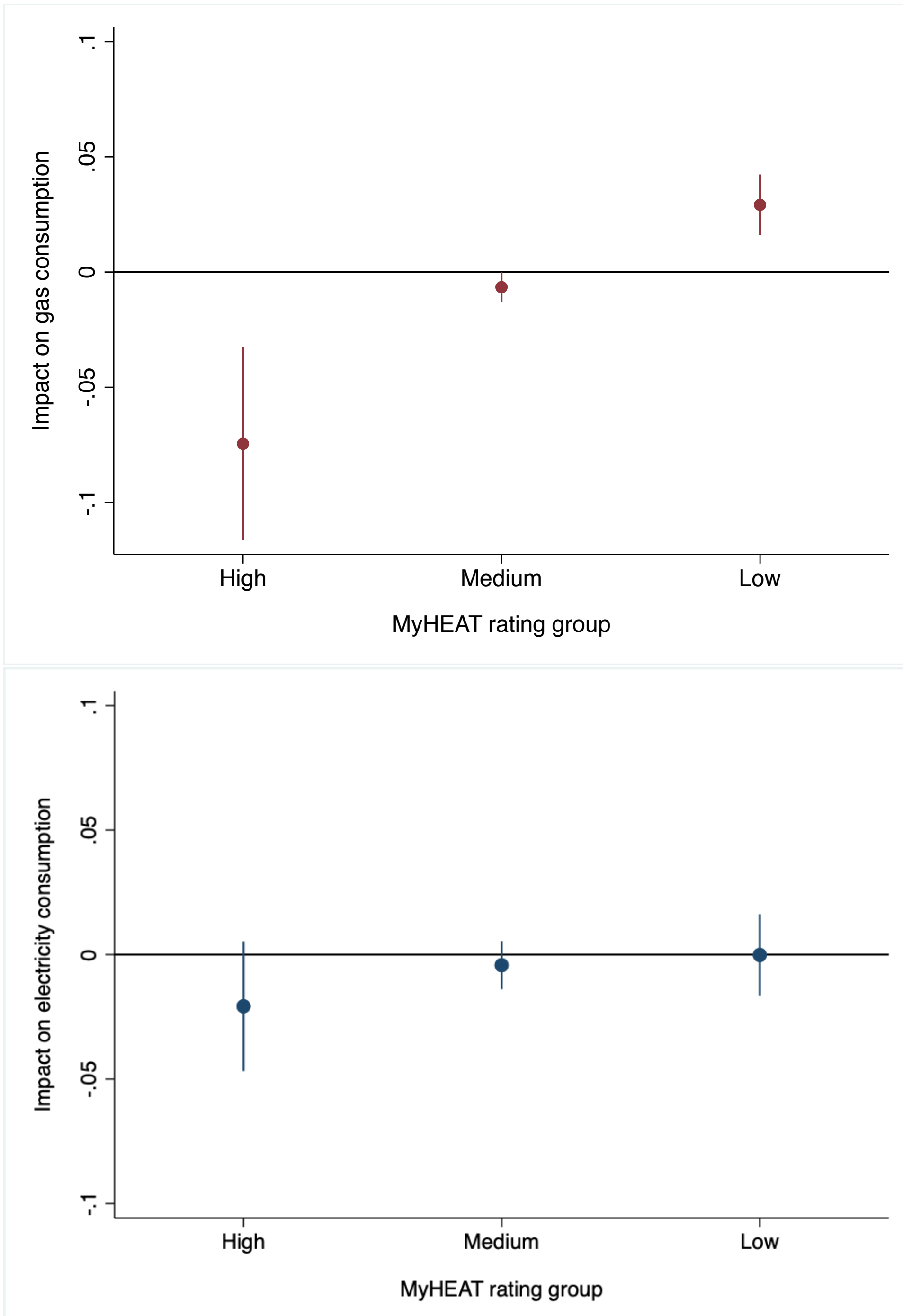
Notes: This figure shows the results of a regression of annual natural gas consumption (left panel) and electricity consumption (right panel) on MyHEAT score, building size, and building type. In each case, the reference category is houses with a MyHEAT rating of 5.

Figure 4: Heterogeneous effects of traditional HER by dollar savings messaging.



Notes: Point estimates of the effect on consumption and confidence intervals for the treatment effect across traditional HER dollar savings messaging. 'Negative' denotes customers who were told they were consuming less and therefore saving on billing expenditures relative to an average similar household; 'Zero' denotes customers who were told they were saving zero dollars relative to similar households; and 'Positive' denotes customers who were consuming more than similar households and therefore told they were paying more money than average.

Figure 5: Heterogeneous effects of MyHEAT treatment according to MyHEAT score.



Notes: Point estimates of the effect on consumption and confidence intervals for the treatment effect across MyHEAT rating groups. 'High' denotes ratings of 8, 9 and 10; 'Medium' denotes ratings of 4,5,6 and 7; and 'low' group denotes ratings of 1,2 and 3.

Table 1: Balance statistics

Experimental Groups and Pre-Treament Balance	Control	Treatment 1	Treatment 2
Electricity (kWh/day)			
Mean	23.93	23.93	23.71
s.d.	15.23	15.17	15.32
t-statistic		0.01	-0.71
Natural gas (mcf/day)			
Mean	0.52	0.51	0.52
s.d.	0.32	0.31	0.32
t-statistic		-0.50	-0.28
Size (m^2)			
Mean	121.83	121.12	121.42
s.d.	42.22	41.75	42.31
t-statistic		-0.38	-0.21
Assessed value			
Mean	298,714	297,931	298,227
s.d.	107,647	108,697	109,651
t-statistic		-0.17	-0.22
Year built			
Mean	1980	1980	1980
s.d.	23	23	24
t-statistic		0.07	-0.56
Heastcore			
Mean	5.21	5.19	5.23
s.d.	1.49	1.50	1.52
t-statistic		-0.54	0.67
Number of Households	4,565	4,642	4,663

Notes: The mean and standard deviation for the consumption and hedonic characteristic variables are presented, along with balance statistics for each treatment group relative to the control. The t-statistic is for a test of the hypothesis that the mean values in the treatment and control group are equal. Electricity and gas consumption are measured in 2017, before the treatment was initiated.

Table 2: Summary statistics for the data

Statistic	N	Mean	St. Dev.	Min	Max
CIVPIN	14,373	60,299.8	91,115.3	26	245,694
locationID	14,373	514,906.6	9,200.9	500,001	534,642
rndgroup	14,373	1.0	0.8	0	2
HEATSCORE	14,373	5.2	1.5	1	10
NEIGHSCORE	14,360	4.7	0.7	3.0	6.0
yearBuilt	14,373	1,979.8	23.3	1,900	2,016
buildingSize	14,373	121.2	42.0	22.0	599.4
AssessValue	14,373	294,482.4	111,749.1	0	1,500,700
Electricity	14,373	8,555.0	4,080.2	4	74,445
Gas	14,373	96.6	37.8	2.9	755.5
Water	14,373	316.6	183.4	1	2,376

Table 3: Estimated relationship between MyHeat rating and natural gas consumption in the pre-treatment period

	Dependent variable:				
	log(Gas)				
	(1)	(2)	(3)	(4)	(5)
HEATSCORE	0.041***	0.047***	0.031***	0.033***	0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Building size	X	X	X	X	X
Building type		X	X	X	X
Year built			X	X	X
Neighbourhood				X	X
Street name				X	X
Assessed value					X
Observations	14,373	14,373	14,373	14,360	14,358
R^2	0.218	0.358	0.406	0.453	0.456
Note:					*p<0.1; **p<0.05; ***p<0.01
					Robust standard errors in parentheses

Table 4: Estimated relationship between MyHeat rating and electricity consumption in the pre-treatment period

	Dependent variable:				
	log(Electricity)				
	(1)	(2)	(3)	(4)	(5)
HEATSCORE	−0.001	0.013***	0.018***	0.017***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Building size	X	X	X	X	X
Building type		X	X	X	X
Year built			X	X	X
Neighbourhood				X	X
Street name				X	X
Assessed value					X
Observations	14,373	14,373	14,360	14,360	14,358
R^2	0.121	0.170	0.185	0.212	0.216
Note:					*p<0.1; **p<0.05; ***p<0.01
					Robust standard errors in parentheses

Table 5: Difference-in-difference regression results for natural gas and electricity

Dependent variable:	Daily Gas Use		Daily Electricity Use	
	(1)	(2)	(3)	(4)
Any treatment	-0.003		-0.002	
T_i x P_{it}	(0.003)		(0.004)	
Traditional Social Comparison		-0.001		-0.004
T_{i1} x P_{it}		(0.003)		(0.005)
MyHeat Social Comparison		-0.006*		0.001
T_{i2} x P_{it}		(0.003)		(0.005)
Observations	3,797,726	3,797,726	3,797,726	3,797,726
R-squared	0.86	0.86	0.69	0.69

Notes: Standard errors are two-way clustered by household and day-of-sample, shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Difference-in-difference regression results with heterogeneous treatment effects for natural gas and electricity

Dependent variable:	Daily Gas Use		Daily Electricity Use	
	(1)	(2)	(3)	(4)
Traditional Social Comparison		0.002		-0.004
$T_{i1} \times P_{it}$		(0.003)		(0.005)
Traditional Social Comparison x Dollar Savings	-0.012***	-0.012***	-0.002	-0.002
$D_{i1m} \times T_{i1} \times P_{it}$	(0.003)	(0.003)	(0.002)	(0.002)
MyHeat Social Comparison		0.105***		0.040***
$T_{i2} \times P_{it}$		(0.022)		(0.013)
MyHeat Social Comparison x Dollar Savings	-0.026***	-0.075***	-0.007**	-0.027***
$D_{i2m} \times T_{i2} \times P_{it}$	(0.006)	(0.016)	(0.004)	(0.009)
Observations	3,797,726	3,797,726	3,797,726	3,797,726
R-squared	0.86	0.86	0.69	0.69

Notes: Standard errors are two-way clustered by household and day-of-sample, shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Impact on energy efficiency program participation

	Dependent variable:			
	Program participation			
	Any program	Home improvement	Online rebates	Home energy plan
	(1)	(2)	(3)	(4)
factor(treatment)1	−0.007	0.001	-0.007^{**}	−0.001
	(0.005)	(0.004)	(0.003)	(0.002)
factor(treatment)2	0.010^{*}	0.007^{**}	0.004	−0.002
	(0.005)	(0.004)	(0.003)	(0.002)
Constant	0.061^{***}	0.029^{***}	0.026^{***}	0.006^{***}
	(0.004)	(0.003)	(0.002)	(0.001)
Observations	13,931	13,931	13,931	13,931
R^2	0.001	0.0003	0.001	0.0001
Adjusted R^2	0.001	0.0002	0.001	−0.00004
Residual Std. Error (df = 13928)	0.241	0.176	0.155	0.074
F Statistic (df = 2; 13928)	5.351^{***}	2.192	6.641^{***}	0.742

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

B Analysis of statistical power

We use a simulation-based approach to determining statistical power, which accounts for the more complex research design that we follow in this study. In particular, the simulation-based approach to power analysis makes it possible to include pre-post measurements for each household, as well as to account for the correlation in measurements of electricity and gas consumption over time within household accounts. The simulation-based approach to determining statistical power uses data on electricity and gas consumption for Medicine Hat residential customers from 2015 and 2016 (preceding the real treatment). In the simulation-based approach to power analysis, we treat 2016 as the "treatment" period and 2015 as the "pre-treatment" period. The simulation proceeds as follows:

1. Choose a sample size, N, which will be divided evenly between control and treatment households.

2. Choose a Type I error rate, α. In the power analysis, $\alpha = 0.05$.

3. Randomly select with replacement $N/2$ treatment households and $N/2$ control households from the 2015-16 consumption data.

4. Pick an effect size, τ, which the the effect of the treatment on treated households in the treated period. For these households in this period, reduce energy (gas, electricity) consumption by τ.

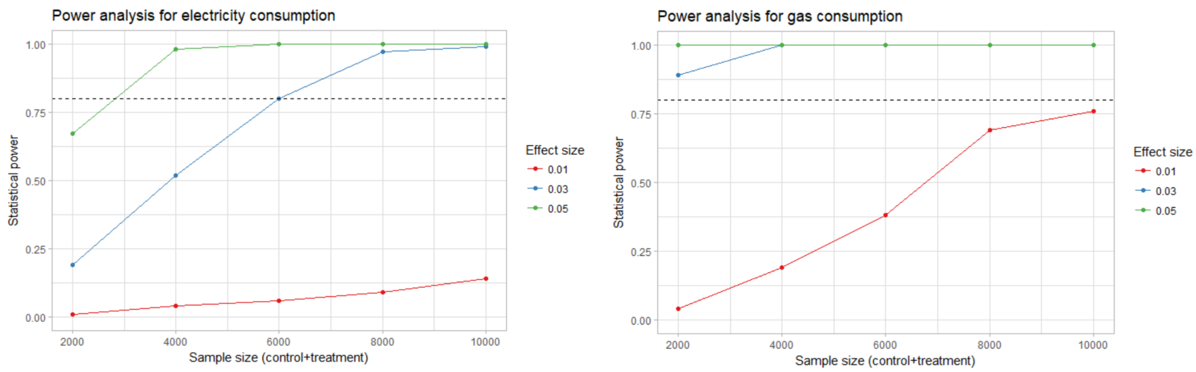
5. Use the data set to estimate the effect of treatment on treated households, following the same estimation strategy and approach to inference as described in the paper (i.e., difference in difference with household and date fixed effects and standard errors clustered on the household).

6. Record whether the coefficient estimate for τ has a p-value below α. If it does, the trial has succeeded in finding an effect where it exists. If it does not, the trial has failed to find an effect where it exists.

7. Repeat the above simulation a large number of times. The ratio of successes to the number of simulations is the statistical power. We repeated the simulation 100 times for each combination of N and τ.

8. Repeat the above simulation with different values for N and τ. We tried values of N from 2,000 to 10,000 and values of τ from 0.01 to 0.05. This gave a total of 15 sets of simulations (100 each) for both gas and electricity.

Figure 6: Results of analysis of statistical power.



Based on this approach, we determine the statistical power of the experiment. Results of the power analysis are given in Figure 6. For electricity, the results suggest that a power of 0.8 is achieved with about 2,500 households for a effect size of 0.05 (i.e., 5% reduction in electricity consumption). Achieving a power of 0.8 for an effect size of 0.03 requires about 6,000 households. With an effect size of 0.01, even 10,000 households is not sufficient to produce a high-powered experiment.

The results for natural gas consumption show that for the same number of households and the same treatment effect, statistical power is substantially higher. This is because the variability is natural gas consumption over time within households is much lower than for electricity. For natural gas, if the treatment effect is 0.03 or larger, an experimental population of 2,000 produces a high-powered experiment. For a treatment effect of 0.01, a sample size of 10,000 produces a statistical power of about 0.8.

Prior results using social comparisons and information feedbacks to energy consumers provide some guidance for the *ex ante* selection of the effect size. For example, Allcott (2011) estimates a treatment effect of about 0.02 based on social comparisons included in electricity bills; Martin and Rivers (2018) estimates a treatment effect of 0.03 based on real-time feedback with in-house displays; and Gleerup et al. (2010) estimates a treatment effect of 0.03 based on consumption feedback using text messages. Although the context and technology for this study differs from those above, these do suggest that a treatment effect of 0.02 to 0.03 is plausible. Based on this assumption, as well as the power analysis above, we design the experiment such that each treatment group and the control group contains roughly 5,000 households (corresponding to $N = 10,000$ in Figure 6). This should provide us with a high power to detect the treatment effect, conditional on its existence.

C Sample utility bills with experimental treatments

C.1 Treatment 1

Medicine Hat
The Gas City

Utility Statement February 14 2018
580 1 St SE, Medicine Hat, AB T1A 8E6
customer_accounts@medicinehat.ca
403 529 8111

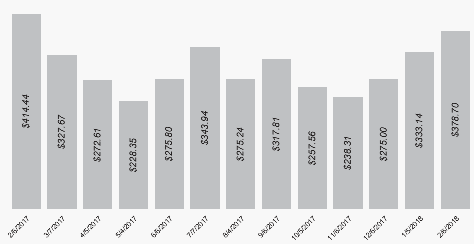
C-10

[redacted]

[redacted] [redacted]

Bill Period **Jan 06 to Feb 06**

Your billed amounts history:



Knowledge Saves Power

Last month you used **30% more energy** than similar homes. Over the course of a year, your **high** consumption could **cost** you **$390**.

See page 4 for your personalized comparison and options to save energy.

You currently owe 378.70
Automatic withdrawal date Mar 8 2018

Your account activity

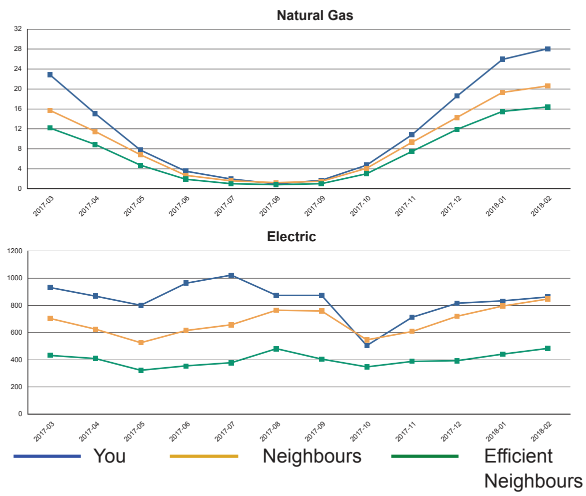
Amount on your last bill	333.14
Payment (Feb 8, 2018)	-333.14
Your balance forward	0.00

Current Charges

*Electric (862 kwh)	99.56
*Gas (28.05 GJ)	169.79
Water (3.00 CM)	30.79
Sewer	42.18
Solid Waste	22.91
*GST(Registration 121408967 RT0001)	13.47
Total new charges	**378.70**
Total you now owe	**378.70**
After March 8 pay	386.27

Automatic withdrawal date Mar 8 2018

You used 30% more energy than similar homes in Medicine Hat last month.



The graphs above show the comparison of your home's energy usage to similar homes in Medicine Hat. Your Neighbours are others in the city with homes that are similar in size and age. Your energy efficient neighbours are those that fall within the lowest 20% of energy users for similar homes in the city.

Based on your energy usage last year, you could end up spending **$390** more per year on your utility bills when compared to your neighbours.

See a breakdown of your home's energy consumption by signing up for the City's eUtility service.

KNOWLEDGE SAVES POWER

What can you do to save?

Seal Air Leaks	You may be eligible for a rebate of up to $700 from HAT Smart for reducing air leakage in your home.
Turn Down the Heat	Avoid heat loss by simply turning down the heat to 16°C when you leave home.

Learn more at **www.hatsmart.ca**

Upgrade Your Insulation	You may be eligible for a rebate of up to $3,500 from Energy Efficiency Alberta for upgrading insulation in your home.
Install New Windows	You may be eligible for a rebate of up to $1,500 from Energy Efficiency Alberta for switching to efficient windows.

Learn more at **www.efficiencyalberta.ca**

For more information on the Knowledge Saves Power project, visit **www.hatsmart.ca** or call **403.502.8799**.

C.2 Treatment 2

Medicine Hat
The Gas City

Utility Statement February 14 2018
580 1 St SE, Medicine Hat, AB T1A 8E6
customer_accounts@medicinehat.ca
403 529 8111

C-10

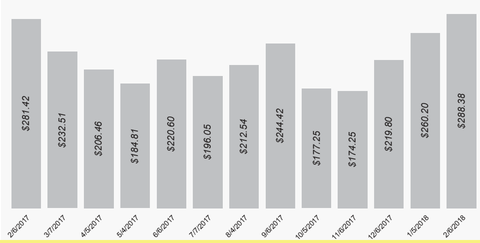
Utility bill for:

[redacted]

[redacted]

Bill Period **Jan 06 to Feb 06**

Your billed amounts history:



Knowledge Saves Power

Your home's heat loss rate is **average.** You could **save $125** per year on your bills by improving this score.

See page 4 for your personalized comparison and options to save energy.

You currently owe 288.38
Please pay by March 13 2018

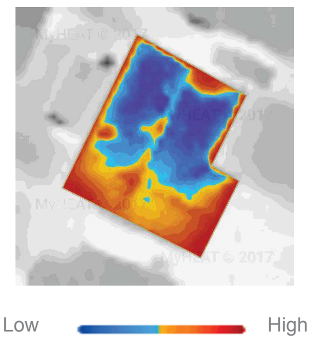
Your account activity

Amount on your last bill	260.20
Payment (Feb 1, 2018)	-260.20
Your balance forward	0.00

Current Charges

*Electric (518 kwh)	64.46
*Gas (18.40 GJ)	112.85
Water (9.00 CM)	37.12
Sewer	42.18
Solid Waste	22.91
*GST(Registration 121408967 RT0001)	8.86
Total new charges	**288.38**
Total you now owe	**288.38**
After March 13 pay	294.15

Your home has a medium heat loss rate with a score of 5/10



The lower the rating, the less heat is leaving your home. You could save **$125** per year on your bills by lowering this score.

The thermal image was taken of your home's roof using an infrared camera in fall 2017. This image can help you identify air leaks that may be wasting energy in your home and resulting in higher bills.

Red areas on your heat map show potential heat loss and can be improved with simple weatherization techniques.

For more information on your home's MyHeat score, visit
www.myheat.ca/thehat/EJMDXA.

KNOWLEDGE SAVES POWER

What can you do to save?

Seal Air Leaks	You may be eligible for a rebate of up to $700 from HAT Smart for reducing air leakage in your home.
Turn Down the Heat	Avoid heat loss by simply turning down the heat to 16°C when you leave home.

Learn more at **www.hatsmart.ca**

Upgrade Your Insulation	You may be eligible for a rebate of up to $3,500 from Energy Efficiency Alberta for upgrading insulation in your home.
Install New Windows	You may be eligible for a rebate of up to $1,500 from Energy Efficiency Alberta for switching to efficient windows.

Learn more at **www.efficiencyalberta.ca**

For more information on the Knowledge Saves Power project,
visit **www.hatsmart.ca** or call **403.502.8799**.