CADMUS

Memorandum

To: Mitch Horrie; Public Service Commission of Wisconsin

From: Yu Wu, Alexa Wallace, and Jill Krueger; Cadmus

Subject: Focus on Energy Save to Give Rural Behavior Pilot Impact Evaluation

Date: May 28, 2024

In calendar year (CY) 2022, Focus on Energy implemented Phase II of the Save to Give Pilot in the communities of New Richmond and Mount Horeb, Wisconsin. Within each community, Focus on Energy fielded three, one-month long campaigns to engage with residents and encourage participation. During campaigns, participants received digital and in-person prompts to take energy-saving actions and record their actions on an online platform to earn donations for a local nonprofit organization. In New Richmond, Campaign One was fielded in January, Campaign Two was fielded in April, and Campaign Three was fielded in October. In Mount Horeb, Campaign One was fielded in March, Campaign Two was fielded in June, and Campaign Three was fielded in October. All campaigns took place in 2022.

The evaluation team conducted an impact evaluation to determine the energy savings achieved during Phase II of the pilot. This memo describes the team's approach to estimating the pilot's effect on participants' energy use during Campaign One, documents evaluation findings, and proposes recommendations.

Executive Summary

As summarized in Table 1, there are two main takeaways from this evaluation:

- Save to Give participants achieved 2.5% electricity savings and -0.3% gas savings in New Richmond and -1% electricity savings and 1.6% gas savings in Mount Horeb. However, only electricity savings in New Richmond are statistically significant.
- The impact of the Save to Give Rural Behavior Pilot on electricity consumption is greater than the impact on gas consumption.

Table 1. Savings Summary

Location	# of Treated Customers	# of Control Customers	Average Daily Savings per Customer	Total Annual Savings	% Savings
New Richmond, Electricity	508	356	0.64 kWh	118,135 kWh	2.5%
New Richmond, Gas	345	306	-0.004 therms	-506 therms	-0.3%
Mt. Horeb, Electricity	184	151	-0.19 kWh	-10,897 kWh	-1.0%
Mt. Horeb, Gas	153	144	0.02 therms	945 therms	1.6%

The pilot ended at the end of CY 2022. Due to high savings acquisition costs, the pilot was not extended nor integrated into existing program offerings. Should Focus on Energy decide to reintroduce the pilot, the evaluation team offers two recommendations based on its evaluation findings:

- To increase statistical significance and likelihood of identifying pilot impacts, increase sample sizes in smaller communities like Mount Horeb by boosting participant recruitment and engagement. Leveraging local partnerships and community events remains crucial. Future efforts should target a minimum enrollment of 1,000 participants.
- To increase the likelihood of reducing natural gas consumption, tailor interventions in a way that directly addresses gas usage patterns, which tend to be more stable than electric usage patterns.

Introduction

Focus on Energy launched the Save to Give Rural Behavior Pilot in 2021 to serve rural residential utility customers in Wisconsin with a unique energy efficiency offering. The pilot used a community-based behavior program design, where customers in select rural communities could sign up to take on energy efficiency challenges. In exchange for completing those challenges, participants earned points toward a monetary donation to a local nonprofit. Focus on Energy selected New Richmond and Mount Horeb to partake in the second year of the pilot (Phase II). New Richmond, with a population of approximately 10,000, is located in St. Croix County in western Wisconsin. Its schools, businesses, and active community make New Richmond a center for surrounding towns. Mount Horeb, home to about 7,500 people, is situated in Dane County, Wisconsin. Mount Horeb is a small, independent community with a strong identity. Focus on Energy requested the evaluation team assess the impact of the second year of the pilot (Phase II) on energy consumption.

Data and Methodology

The evaluation team collected billing data from utilities and participation data from implementers in New Richmond and Mount Horeb. Billing data spanned a year before the date of Campaign One in each treatment group through one year after the campaign date to allow for 12 months of pre-treatment data and 12 months of post-treatment data.[1] The evaluation team also collected billing data from a representative control group for each treatment community, Hudson near New Richmond, and Sauk City near Mount Horeb. Billing data contained customer name, address, meter read dates, and monthly usage by fuel type.

The team obtained daily average temperature data for the pre- and post-treatment periods from National Oceanic and Atmospheric Administration (NOAA) weather stations representing all ZIP codes

[1] Pre- and post-treatment periods were determined by Campaign One dates. In New Richmond, Campaign One started on January 1, 2022. The pre-treatment period is January 1, 2021 to December 31, 2021, and the post-treatment period is January 1, 2022 to December 31, 2023. In Mount Horeb, Campaign One started on March 1, 2022. The pre-treatment period is March 1, 2021 to February 28, 2022, and the post-treatment period is March 1, 2022 to February 28, 2023.

associated with the participants and control group. The team also used the Typical Meteorological Year, version 3 (TMY3) Normal series (1991-2005) to weather-normalize the billing data.

The team cleaned the data from New Richmond and its control group (Hudson) separately from Mount Horeb and its control group (Sauk City) as well as by fuel type. The team followed the below steps to prepare the data for analysis:

1. Clean participant tracking information and identify eligible participant population (customers with active accounts during the analysis period).
2. Clean billing data to remove any data anomalies in the billing data (i.e., negative electric or gas reads, duplicated bills, and insufficient billing data with less than 25 days between meter reads or extended read dates across multiple months).
3. Map zip codes to the nearest weather station, obtain daily average temperatures over pre- and post-treatment periods, and calculate total heating degree days (HDDs) and cooling degree days (CDDs) using the base temperature of 65°F occurring during each billing period.
4. Calendarize billing and weather data. The team allocated energy use, HDDs, and CDDs occurring during each billing cycle to calendar months based on the number of days in the billing cycle falling in each calendar month.

Matching Methods for Selecting Control Groups

The Save to Give Pilot functions as an opt-in program, meaning customers enroll at their own discretion rather than being automatically enrolled. While the evaluation team selected two control communities corresponding to the two treatment communities to evaluate the pilot, it did not define the specific control groups prior to the pilot's start. To develop the control groups, the team leveraged the population of residential billing data in control communities and performed matching including Propensity Score Matching and Mahalanobis Metric Matching.[2]

In Propensity Score Matching, the team estimated the probability—or propensity—of participating in the Save to Give Pilot (in the treatment group) as a function of seasonal average daily energy consumption or monthly average daily energy consumption in the year preceding the launch of the Save to Give Campaign Two in January 2022 (for New Richmond) and March 2022 (for Mount Horeb). The team matched each Save to Give participant to one nonparticipant in the control community whose propensity score was closest to that of the participant (i.e., using nearest-neighbor matching).

To ensure best-balanced matches, the team also employed Mahalanobis Metric Matching. This method calculates the Mahalanobis distance between each treatment group customers and control group

---

[2] Given that the treatment is not randomly assigned, a quasi-experimental method must be employed to estimate the causal impact. A challenge to quasi-experimental studies is the occurrence of selection bias, which may lead to a loss of internal validity. Propensity score matching aims to reduce selection bias by matching on observable characteristics that could otherwise confound estimation of the treatment effect.

Abadie, Alberto, and Guido W. Imbens. 2016. "Matching on the Estimated Propensity Score." *Econometrica* (84, no. 2): 781–807.

customers using seasonal average daily energy consumption or monthly average daily energy consumption in the 12 months prior to the Save to Give Campaign One. Unlike Propensity Score Matching, Mahalanobis Metric Matching uses the covariance among these variables to more accurately reflect the multidimensional similarity between individuals.

Since the team expected that New Richmond participants and Mount Horeb participants would differ both in their typical energy consumption patterns and their eventual Save to Give savings, it separately conducted matching for the two communities' participants within each fuel type.

To ensure high-quality matching, the team conducted statistical equivalence tests to confirm that each matched control group had similar seasonal consumption levels to those of Save to Give participants before their first campaign advertising date. Ideally, there should be no statistically significant differences in annual or within-season electricity consumption between the groups.

Based on the matching results, the team selected either Propensity Score Matching or Mahalanobis Metric Matching for the final model, using the method that achieved the best balance between the treatment and control groups.

Billing Analysis

The evaluation team used panel regression models to analyze the effect of the Save of Give treatment on energy consumption. The analyses conformed to the approach described in the Uniform Methods Project.[3] The team used both post-only models and difference-in-differences models and tested different regression model specifications.[4] The models controlled for pre-Save to Give energy consumption patterns, customer-specific fixed effects (average consumption for each unique customer ID), time fixed effects (average consumption for each calendar month), and weather effects (HDDs and CDDs calculated at 65°F base temperature). The team clustered standard errors on each pair of participants and matched nonparticipants. Consistent with the matching, the team estimated separate models for New Richmond and Mount Horeb customers, and for electric customers and natural gas customers.

---

[3] Stewart, Jim, and A. Todd. 2020. *The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures*. "Chapter 17: Residential Behavior Protocol." Prepared for National Renewable Energy Laboratory, Golden, Colorado. NREL/SR-7A40-77435. https://www.nrel.gov/docs/fy21osti/77435.pdf

[4] Post-only model regresses each customer's average daily energy consumption (in a given month) on a treatment indicator variable and other control variables. The control variables (regressors) include the customers' pre-treatment energy use by month, month-by-year fixed effects, and weather.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* (104, no. 10): 3003–3037.

Savings Calculation

The evaluation team estimated the percentage of net savings, which is the ratio of average daily savings per customer to the control group's average daily consumption in the post-participation period. These savings serve as an estimate of the baseline energy use absent the pilot.

The team also estimated annual net savings as the product of average daily savings per participant and the total number of days all participants were treated, referred to as treatment days. If a customer was active for the whole year (a full year starting from the date of Campaign One), the number of treatment days was 365. If a customer was inactive for part of the year, the number of treatment days was the number of days from the start date of treatment to the move-out date.

Preliminary Results

The evaluation team conducted separate analyses for New Richmond and Mount Horeb. Those results are summarized below.

New Richmond

The evaluation team analyzed both electric and gas billing data for New Richmond.

Matching for Electric Customers

The evaluation team used the Mahalanobis distance to match participants, incorporating monthly average daily consumption (ADC) in the pre-treatment period as a key covariate.[5] To assess the quality of the resulting matches, the team examined the covariate balance after matching (Table 2).[6] Standardized mean differences are close to zero and variance ratios are close to 1, indicating good balance. In addition, based on a two-sample t-test, the mean of pre-treatment period monthly ADC in the treatment group is not statistically different from the mean of pre-treatment period monthly ADC in the control group.

---

[5] The evaluation team also tested to include winter ADC, summer ADC, shoulder ADC, and total average ADC in the pre-treatment periods in the calculation of Mahalanobis distance. Additionally, the team conducted Propensity Scores Matching. The final model is one that resulted in best fitness of model (lowest AIC) and matching balance.

[6] Assessing balance involves assessing whether the distributions of covariates are similar between the treated and control groups. Balance is typically assessed by examining univariate balance summary statistics for each covariate.

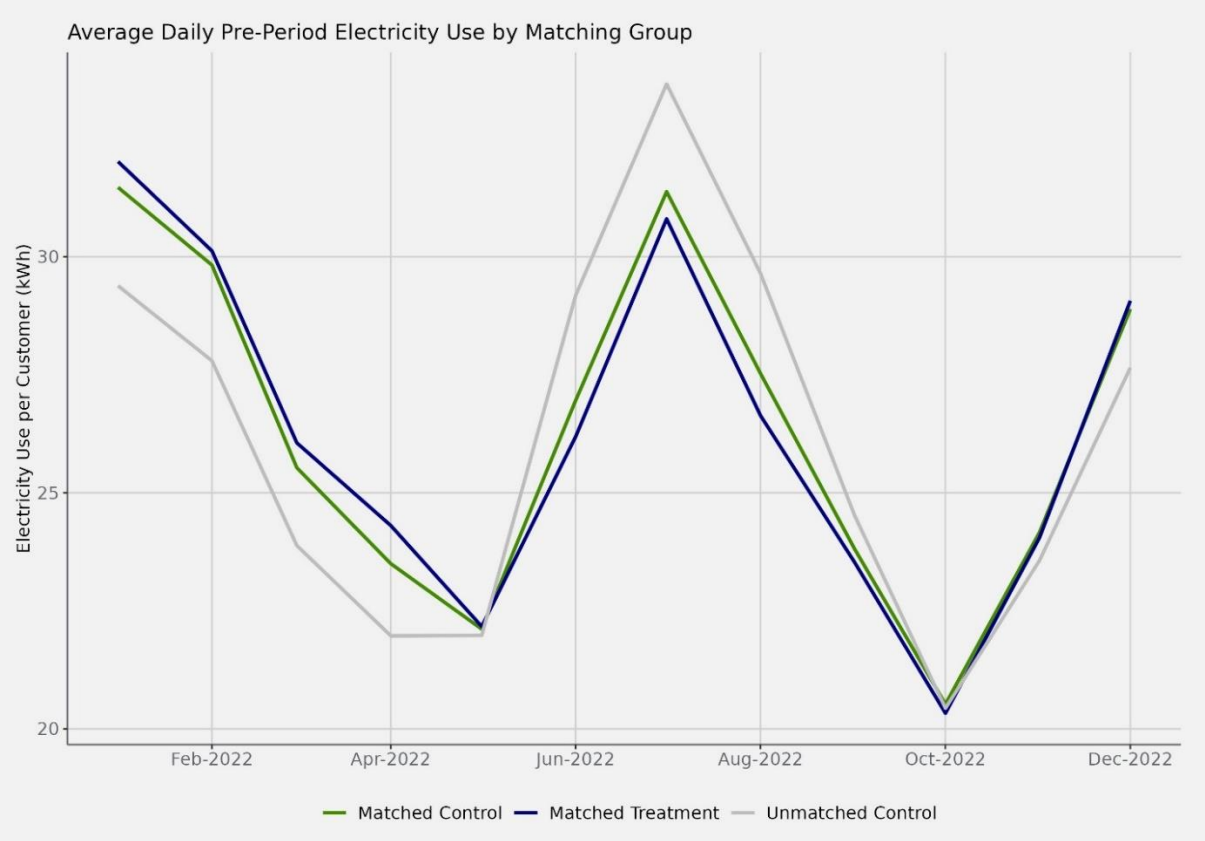
Greifer, N. 2021. *Assessing balance*. The R foundation. https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html

Table 2. Summary of Balance after Matching for New Richmond Electric Customers

	Means Treated	Means Control	Std. Mean Differences	Variance Ratio	Pair Distances
Pre-period Feb. ADC	30.78	28.75	0.08	1.39	0.28
Pre-period Mar. ADC	26.57	24.89	0.09	1.30	0.29
Pre-period Apr. ADC	24.56	23.10	0.10	1.27	0.31
Pre-period May. ADC	22.71	21.74	0.08	1.13	0.39
Pre-period Jun. ADC	26.82	26.36	0.04	1.05	0.45
Pre-period Jul. ADC	31.22	30.61	0.04	1.10	0.44
Pre-period Aug. ADC	27.44	27.03	0.03	1.06	0.45
Pre-period Sep. ADC	23.95	23.30	0.05	1.08	0.42
Pre-period Oct. ADC	20.65	19.98	0.06	1.28	0.36
Pre-period Nov. ADC	24.41	23.47	0.06	1.33	0.33
Pre-period Dec. ADC	29.49	27.85	0.08	1.38	0.31

In addition, the team conducted visual diagnostics to look for differences between the groups. Figure 1 shows the monthly electric ADC in the pre-treatment period by matching groups. The treatment group's pre-treatment period ADC is better matched to the control group's pre-treatment period ADC between October to December.

Figure 1. Pre-Treatment Period Average Daily Electricity Consumption in New Richmond



Billing Analysis for Electric Customers

The evaluation team found the treatment effect of the Save to Give Pilot on electricity consumption is statistically significantly negative, meaning participants' consumption was lower in the treatment group than in the control group. The results indicate electricity savings from the pilot. Table 3 shows the regression results of the post-only model, which typically have higher precision and are more robust.[7] The treatment effect is -0.637, meaning average daily savings per customer is 0.637 kWh per day. Using the savings calculation formula referenced in the methodology section, this represents a net percentage savings of approximately 2.5% and total annual savings of 118,135 kWh across all New Richmond participants.

Table 3. Estimation of the Treatment Effects for Electric Customers in New Richmond

	Post-Only Model
Treatment Effect on Electricity Consumption	**-0.637[a]**
	(0.349)
HDD 65	-0.115
	(0.090)
CDD 65	1.234[b]
	(0.226)
Month Fixed Effects × Monthly Pre-Period ADC	Yes
Month Fixed Effects	Yes
Observations	8,148
Adjusted R^2	0.804[c]

Note: This table reports the regression analysis results. The negative value of the treatment effect on energy consumption indicates energy savings. Cluster-robust standard errors are in the parentheses.

[a] These savings were statistically significant at the 90% confidence level.

[b] The explanatory variable is statistically significant at the 99% confidence level.

[c] A higher Adjusted R^2 value (close to 1) indicates a better fit of the model, after accounting for the number of explanatory variables.

Matching for Gas Customers

The evaluation team used the Mahalanobis distance to match participants, incorporating summer ADC in the pre-treatment period as a key covariate.[8] To assess the quality of the resulting matches, the team examined the covariate balance after matching (Table 4). Standardized mean differences are close to zero and variance ratios are close to 1, indicating good balance. Based on a two-sample t-test, the mean of pre-treatment period summer ADC in the treatment group is not statistically different from the mean of pre-treatment period summer ADC in the control group.

---

7 Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* (104, no. 10): 3003–3037.

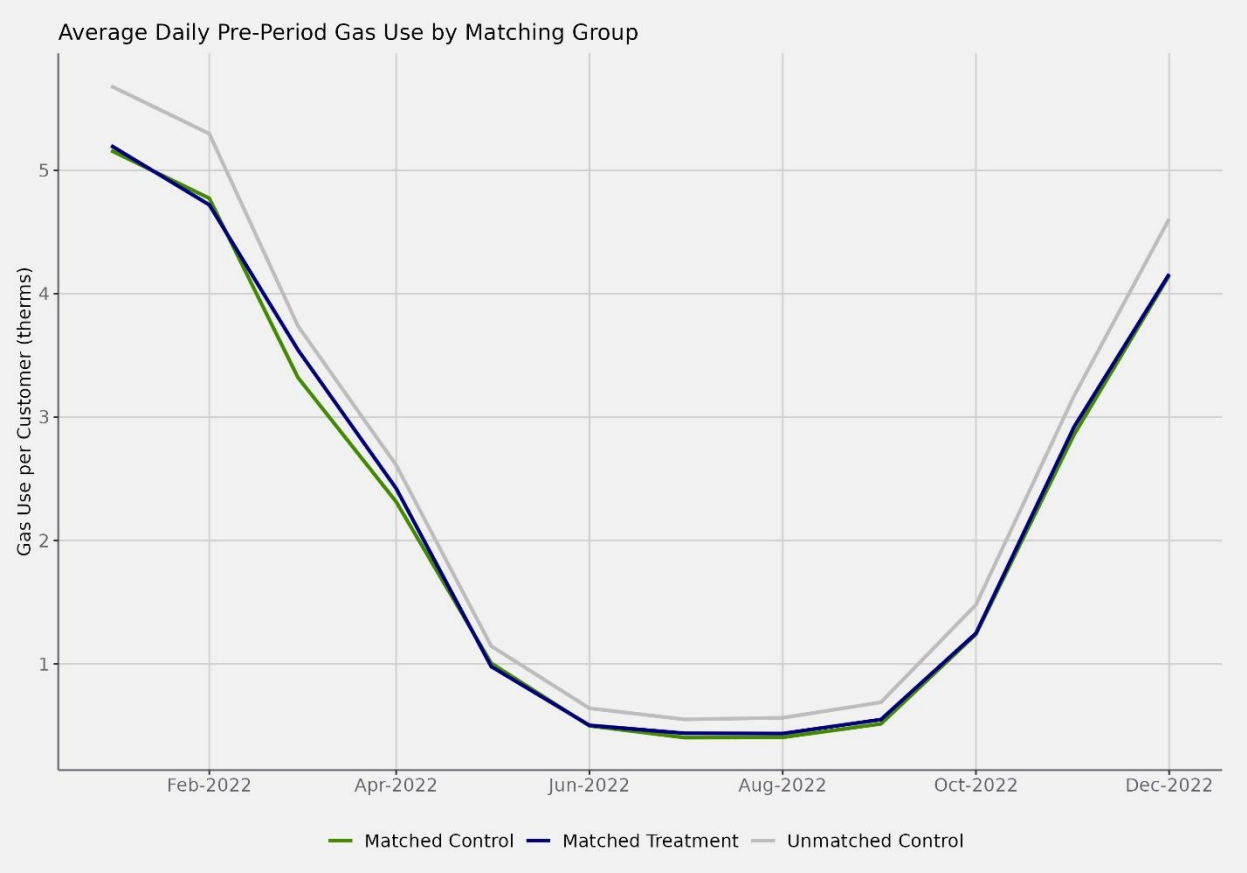
8 The evaluation team also tested to include winter ADC, shoulder ADC, and total average ADC, and each of the 12 months ADC in the pre-treatment periods in the calculation of Mahalanobis distance. Additionally, the team conducted Propensity Scores Matching. The final model is one that resulted in best fitness of model (lowest AIC) and matching balance.

Table 4. Summary of Balance after Matching for New Richmond Gas Customers

	Means Treated	Means Control	Std. Mean Differences	Variance Ratio	Pair Distances
Pre-period Summer ADC	0.40	0.40	0.00	1.00	0.00

In addition, the team conducted visual diagnostics to look for differences between the groups. Figure 2 shows the monthly gas ADC in the pre-treatment period by matching groups. The treatment group's pre-treatment period ADC is better matched to the control group's pre-treatment period ADC during both winter and summer months.

Figure 2. Pre-Treatment Period Average Daily Gas Consumption in New Richmond



Billing Analysis for Gas Customers

The evaluation team found the treatment effect of the Save to Give Pilot on gas consumption is positive, meaning participants' consumption was higher in the treatment group than in the control group with no energy savings. Table 5 shows the regression results of the post-only model. The treatment effect is 0.004, that is average daily savings per customer is -0.004 therms. This represents a net percentage savings of approximately -0.3% and total annual savings of -506 therms across all New Richmond participants.[9] However, the treatment effects and savings are not statistically significant, and the results are not consistent across models.

Table 5. Estimation of the Treatment Effects for Gas Customers in New Richmond

	Post-Only Model
Treatment Effect on Gas Consumption	0.004
	(0.018)
HDD 65	0.063[a]
	(0.009)
CDD 65	-0.018
	(0.013)
Month Fixed Effects × Monthly Pre-Period ADC	Yes
Month Fixed Effects	Yes
Observations	5,918
Adjusted R^2	0.954[b]

Note: This table reports the regression analysis results. The negative value of the treatment effect on energy consumption indicates energy savings. Cluster-robust standard errors are in the parentheses.

[a] The explanatory variable is statistically significant at the 99% confidence level.

[b] A higher Adjusted R^2 value (close to 1) indicates a better fit of the model, after accounting for the number of explanatory variables.

Mount Horeb

The evaluation team analyzed both electric and gas billing data for Mount Horeb.

Matching for Electric Customers

The evaluation team utilized the Mahalanobis distance to match participants, incorporating monthly ADC in the pre-treatment period as a key covariate.[10] Based on a two-sample t-test, the mean of pre-treatment period monthly ADC in the treatment group is not statistically different from the mean of pre-treatment period monthly ADC in the control group.

Table 6 shows the covariates balance after matching. Most of the standardized mean differences are close to zero (absolute values are less than 0.1) and variance ratios are close to 1, indicating good

---

9 The post-control group ADC is 2.21 therms, and the total number of treatment days is 9,441.

10 The evaluation team also tested to include winter ADC, summer ADC, shoulder ADC, and total average ADC in the pre-treatment periods in the calculation of Mahalanobis distance. Additionally, the team conducted Propensity Scores Matching. The final model is one that resulted in best fitness of model (lowest AIC) and matching balance.

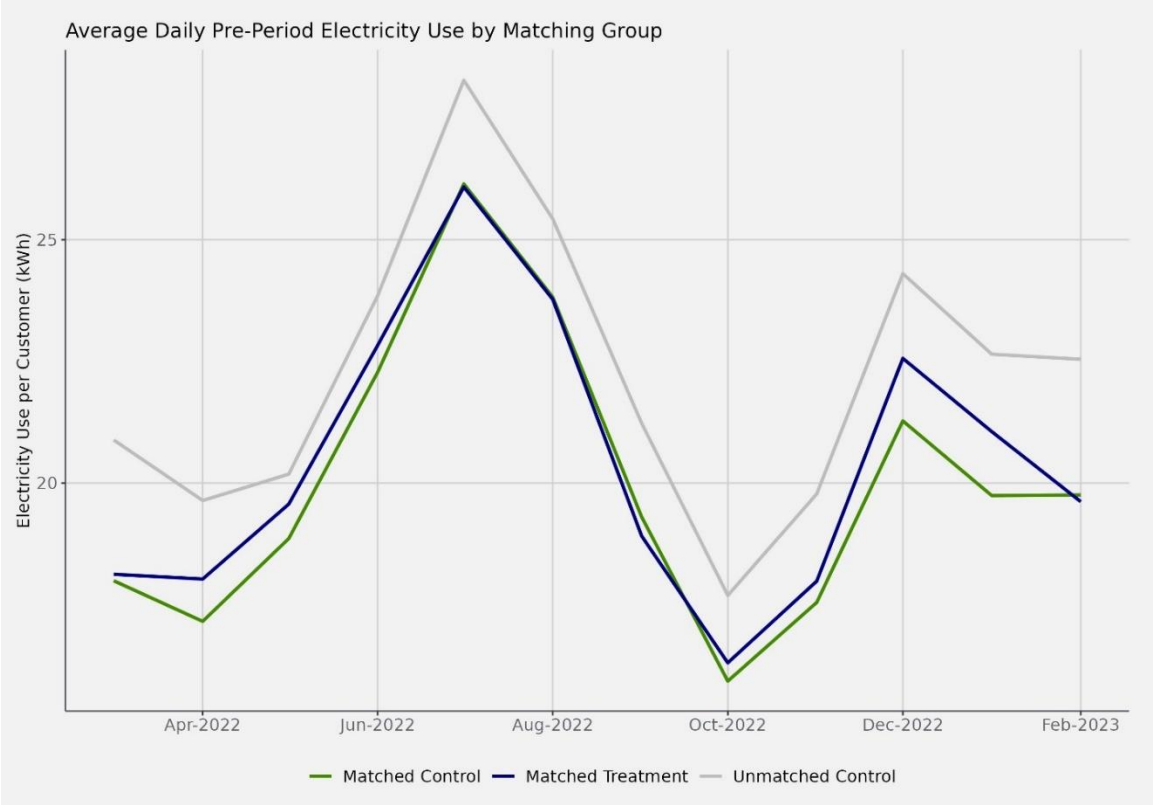
balance in general. Based on a two-sample t-test, the mean of pre-treatment period monthly ADC in the treatment group is not statistically different from the mean of pre-treatment period monthly ADC in the control group.

Table 6. Summary of Balance after Matching for Mount Horeb Electric Customers

	Means Treated	Means Control	Std. Mean Difference	Var. Ratio	Pair Distances
Pre-period Feb. ADC	19.73	19.08	0.05	1.28	0.40
Pre-period Mar. ADC	17.92	17.33	0.05	1.20	0.43
Pre-period Apr. ADC	17.03	16.50	0.05	1.09	0.41
Pre-period May. ADC	19.06	18.15	0.08	1.09	0.44
Pre-period Jun. ADC	22.71	21.68	0.08	1.14	0.42
Pre-period Jul. ADC	26.06	25.41	0.04	1.19	0.39
Pre-period Aug. ADC	23.90	23.06	0.06	1.26	0.39
Pre-period Sep. ADC	19.13	18.65	0.04	1.08	0.44
Pre-period Oct. ADC	15.91	15.45	0.05	1.09	0.44
Pre-period Nov. ADC	17.71	17.01	0.07	1.20	0.44

Visually, the treatment group's pre-treatment period ADC is better matched to the control group's pre-treatment period ADC between July to September (Figure 3). Compared to the unmatched control group customers, the matched control group customers have more similar pre-treatment period energy consumption patterns to the treated customers.

Figure 3. Pre-Treatment Period Average Daily Electricity Consumption in Mount Horeb



Billing Analysis for Electric Customers

The evaluation team found the treatment effect of the Save to Give Pilot on electricity consumption is positive, suggesting higher consumption in the treatment group. However, the treatment effects are not statistically significant. Table 7 shows the regression results of the post-only model. The treatment effect is 0.195, which means negative an average daily savings per customer of -0.195 kWh. Based on the savings calculation formula referenced in the methodology section, the net percentage savings is approximately -1%, and total negative annual savings are -10,897 kWh across all Mount Horeb participants.

Table 7. Estimation of the Treatment Effects for Electric Customers in Mount Horeb

	Post-Only Model
Treatment Effect on Electricity Consumption	0.195
	(0.469)
HDD 65	-0.472
	(0.518)
CDD 65	0.298
	(0.764)
Month Fixed Effects × Monthly Pre-Period ADC	Yes
Month Fixed Effects	Yes
Observations	2,630
Adjusted R^2	0.785[a]

Note: This table reports the regression analysis results. The negative value of the treatment effect on energy consumption indicates energy savings. Cluster-robust standard errors are in the parentheses.

[a] A higher Adjusted R^2 value (close to 1) indicates a better fit of the model, after accounting for the number of explanatory variables.

Matching for Natural Gas Customers

The evaluation team utilized the Mahalanobis distance to match participants, incorporating winter ADC and total ADC in the pre-treatment period as a key covariate.[11] Based on a two-sample t-test, the mean of pre-treatment period winter and total ADC in the treatment group is not statistically different from the mean of pre-treatment period winter and total ADC in the control group.

Table 8 shows the covariates balance after matching. Most of the standardized mean differences are close to zero (absolute values are around 0.1) and variance ratios are close to 1, indicating good balance in general. Based on a two-sample t-test, the mean of pre-treatment period winter and total ADC in the treatment group is not statistically different from the mean of pre-treatment period winter and total ADC in the control group.

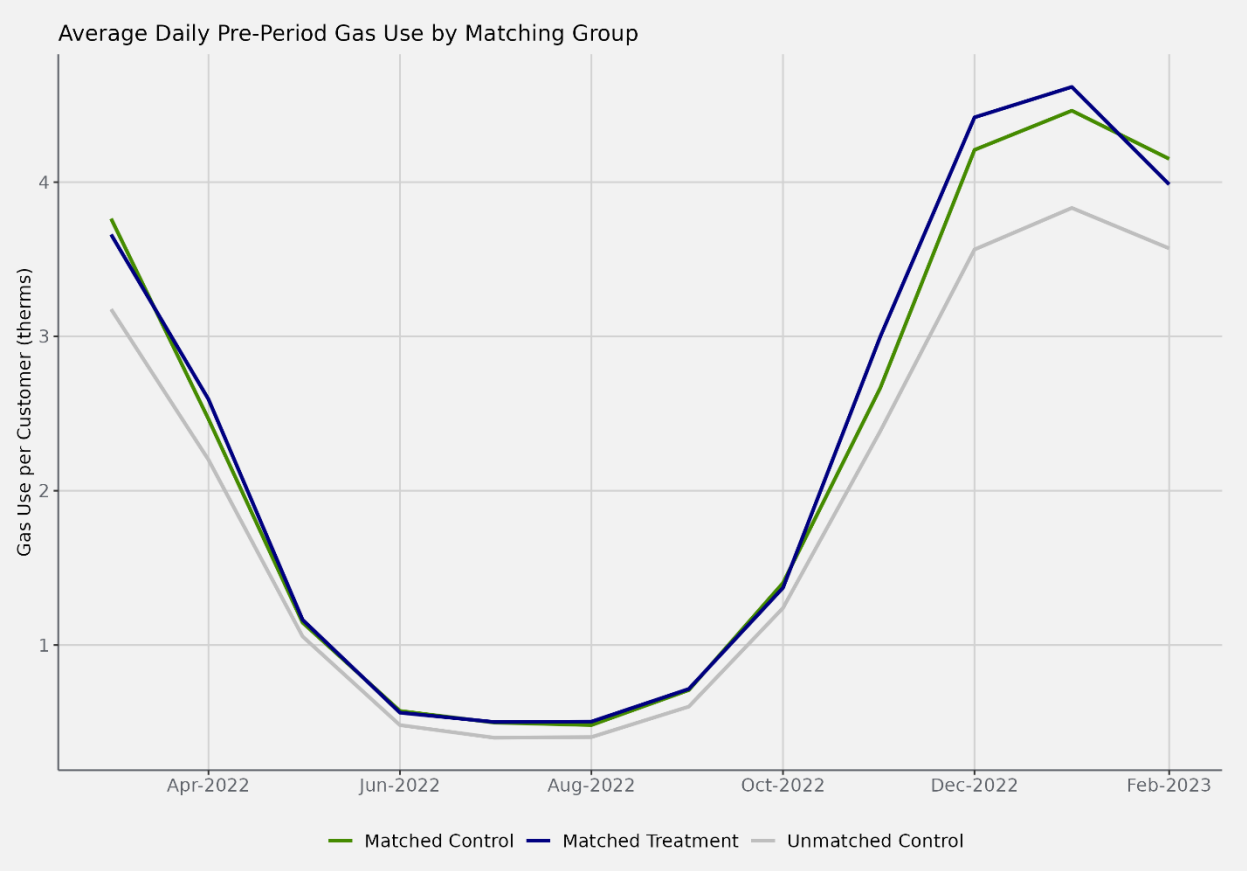
[11] The evaluation team also tested to include summer ADC, shoulder ADC, and each of the 12 months ADC in the pre-treatment periods in the calculation of Mahalanobis distance. Additionally, the team conducted Propensity Scores Matching. The final model is one that resulted in best fitness of model (lowest AIC) and matching balance.

Table 8. Summary of Balance after Matching for Mount Horeb Natural Gas Customers

	Means Treated	Means Control	Std. Mean Difference	Var. Ratio	Pair Distances
Pre-period Winter ADC	4.20	4.20	0.00	1.01	0.04
Pre-period Total ADC	2.23	2.23	0.01	1.03	0.04

Visually, the treatment group's average pre-treatment period ADC is better matched to the control group's pre-treatment period ADC during the summer months (Figure 4). When compared to the unmatched control group customers, the matched control group customers have more similar pre-treatment period energy consumption patterns to the treated customers.

Figure 4. Pre-Treatment Period Average Daily Natural Gas Consumption in Mount Horeb



Billing Analysis for Natural Gas Customers

The evaluation team found the treatment effect of the Save to Give Pilot on gas consumption is negative, meaning consumption was lower in the treatment group than in the control group. However, the treatment effects are not statistically significant. The treatment effect on gas consumption is smaller than for electricity in magnitude, -0.022 (Table 9 on the next page). The net percentage savings is 1.6%, and total annual positive savings are 945 therms across all Mount Horeb participants.

Table 9. Estimation of the Treatment Effects for Natural Gas Customers in Mount Horeb

	Post-Only Model
Treatment Effect on Gas Consumption	**-0.022**
	(0.020)
HDD 65	0.064 [a]
	(0.007)
CDD 65	-0.103 [a]
	(0.026)
Month Fixed Effects × Monthly Pre-Period ADC	Yes
Month Fixed Effects	Yes
Observations	2,207
Adjusted R^2	0.946

Note: This table reports the regression analysis results. The negative value of the treatment effect on energy consumption indicates energy savings. Cluster-robust standard errors are in the parentheses.

[a] The explanatory variable is statistically significant at the 99% confidence level.

[b] A higher Adjusted R^2 value (close to 1) indicates a better fit of the model, after accounting for the number of explanatory variables.

Conclusions and Recommendations

The evaluation team's analysis of Focus on Energy's Save to Give Rural Behavior Pilot Phase II indicated the pilot produced statistically significant electric savings in New Richmond. However, the team observed no statistically significant savings among gas customers in New Richmond or among customers using either type of fuel in Mount Horeb.

Conclusion: The negative savings and statistically insignificant results for gas customers in both communities may stem from behavioral differences in gas versus electricity usage. Unlike electricity, which powers a wide range of devices and can see usage fluctuations throughout the day, gas consumption is often concentrated in specific activities such as heating and cooking. This difference in usage patterns means interventions successful in reducing electricity consumption may not directly translate to gas savings.

Recommendation: Consider future initiatives that more effectively target gas-saving behaviors, possibly through more customized interventions.

Conclusion: The statistically insignificant results observed in Mount Horeb could be attributed to the limitations posed by a small population size and a correspondingly small number of participants. Small sample sizes can significantly hamper the statistical power of a billing analysis, making it difficult to detect true effects or differences if they exist. In such cases, even if the intervention has a positive effect, the small sample size might not provide enough evidence to conclusively demonstrate its impact.

Recommendation: For future initiatives in communities with small populations, consider strategies to enhance participant engagement and increase the sample size. This could involve extended recruitment efforts, broader community engagement and partnership building to raise awareness and interest in the program, or expanding the geographic area of eligible participants.