Incentivising households to reduce their electricity consumption: A meta-analysis of recent experimental evidence

Penelope Buckley*[1]

[1]Univ. Grenoble Alpes, CNRS, INRA, Grenoble INP, GAEL, 38000 Grenoble, France

29th August 2018

Abstract

The present paper employs a meta-analysis approach to analyse the results of recent field experiments and pilot studies which explore the effects of different methods of incentivising electricity consumption reduction on residential consumers' electricity demand. Both data from peer-reviewed research and from grey literature; utility or government reports are used. The strategies currently used in the experimental literature fall into one of two categories: financial incentives (dynamic pricing, monetary information), and non-financial incentives: informational incentives (personal feedback, real-time information) and 'nudges' (social norms, social approval). Previous meta-analyses have reviewed studies from the 1970s and 1980s onwards and conclude that feedback which can be immediately related to the electricity consuming activity, as well as tailored advice and electricity conservation tips are most effective. By focussing only on recent studies (2005 onwards), as well including results from the grey literature, the present paper provides an analysis of studies carried out when more advanced technology has been used and when there has been a greater

*Corresponding author: penelope.buckley@univ-grenoble-alpes.fr
1241 Rue des Résidences - Domaine Universitaire - 38400 Saint-Martin-d'Hères

understanding of the risks of climate change. The analysis includes 105 treatment observations from 39 papers. Results show that, on average, across studies, real-time feedback and monetary information have the greatest effect at reducing electricity consumption. While the effects are slightly smaller, social norms have a significant effect on reducing residential electricity consumption. Compared to previous meta-analysis, the results show that recent studies use larger samples and are generally of high quality (include a control group, subjects are assigned randomly to treatments, demographics and weather are controlled for). As a result, the treatment effects observed are generally smaller than those reported in previous meta-analyses.

Keywords: conservation, consumption, electricity, feedback, incentives, nudges, residential.

Contents

1 Introduction

Up until recently, households were passive consumers of electricity. For the average household, electricity is invisible and consumers are unsure of how much electricity their daily activities consume (Darby et al., 2006; Burgess and Nye, 2008; Hargreaves et al., 2010). However, the electricity grid is undergoing rapid change and is becoming smart. The smart grid allows for improvements in the efficiency, the reliability, and the safety of electricity power infrastructure. It will allow for renewable and other alternative sources of energy to be easily integrated into the grid(Gungor et al., 2011). In addition the new grid allows for the integration of smart enabling technology which will allow residential consumers to take a more active role in their electricity consumption. In turn, this allows households to take action to lower their electricity consumption.

Information on electricity consumption can be collected in varying degrees of granularity and can be communicated to households via various methods. There are currently two principal approaches to lowering electricity demand: (1) more efficient technologies, (2) incentives for consumers to modify their electricity consuming behaviour. Consumers have long been told of the need to turn off unused lights, to not leave appliances on standby, to not leave the fridge door open, to name a few, however, it is now, with the improvements being made to the grid that consumers can receive more accurate information on their demand, and appropriate incentives to lower their electricity consumption.

By providing consumers with information on their previous electricity consumption, informing them of the consequences of increased consumption, and providing them with appropriate incentives, it is predicted that consumers will make the decision to lower their electricity demand (Frederiks et al., 2015). in reality, people do not behave as economic theory would predict. They are subject to biases which affect their ability to assimilate all of the available information in order to do the necessary calculations required to make rational decisions devoid of emotional influences (Thaler and Sunstein, 2008). Given this view of consumers, additional methods of incentivising consumers to lower their electricity consumption which exploit such biases have been tested.

This paper uses a meta-analysis approach to explore the results of contemporary field experiments and pilot studies which have been designed to incentivise consumers to reduce their electricity consumption. The objective is to combine the results of many studies to have a better estimate of the true effect of different types of incentive on residential electricity consumption.

The current meta-analysis adds to literature on meta-analyses which explore incentives for reducing household electricity consumption by including solely recent studies, those published from 2005 up to 2016, the time of data collection. By focussing on this time period, named the "Smart Grid Era" (McKerracher and Torriti, 2013), a more accurate estimate of the effect of an incentive on electricity consumption is calculated. Additionally, the present analysis includes studies from both peer reviewed literature and utility and government reports. Often, previous analyses take into consideration only peer reviewed articles which may result in bias in the selection of studies used for the meta-analysis. The final added-value of the present meta-analysis is the inclusion of a greater level of study design variables, such as how households are recruited to the study, and how they are assigned to the treatment groups.

The following section describes the different incentives used in the experimental literature thus setting out the hypotheses which will be tested, followed by a discussion of previous analyses and reviews in section 3. Section 4 describes the data collection method, the model used and the variables of interest. Section 5 presents the results and finally, section 6 discusses the results and concludes.

2 Incentives for lowering residential electricity consumption

The principal strategies employed to incentivise households to reduce their consumption can be separated into monetary and non-monetary incentives. In this section, we describe the different strategies used in the literature collected in this meta-analysis and thus develop the hypotheses that will be tested.

2.1 Monetary Incentives

Monetary incentives can be separated into one of two categories: electricity cost information and dynamic pricing. We include monetary information here as although it is not a direct monetary incentive, such incentives display information in monetary terms thus informing households of how much they are spending on electricity or how much they are saving. By providing households with information as to how much their electricity consumption costs (as opposed to information on the amount of electricity consumed), households can see the monetary benefits of reducing their electricity consumption. In interviews with households participating in electricity conservation field experiments, residents preferred to receive feedback in monetary terms as this is considered to be more relatable, and more comparable, than energy units (Hargreaves et al., 2010; Raw and Ross, 2011).

Further, with the installation of smart meters in residential homes, a major technological barrier to the implementation of dynamic pricing tariffs has been lifted. Dynamic pricing provides consumers with economic incentives to reduce, or to increase, their electricity consumption by better aligning the retail price of electricity with the wholesale price in order to maintain supply and demand balance in the electricity market (Borenstein et al., 2002). Such pricing tariffs are seen to be effective at reducing demand during periods of high demand but less effective at reducing overall demand (Allcott, 2011; Torriti, 2012).

Hypothesis 1a: *Monetary incentives have negative effect on electricity demand.*
Hypothesis 1b: *Monetary information has a negative effect on electricity demand.*

2.2 Non-monetary Incentives

Non-monetary strategies refer to those which provide households with more detailed information on their electricity consumption. In the experimental literature, this type of incentive can be categorised into personal feedback and social feedback.

2.2.1 Personal Feedback

Personal feedback provides households with data on their own electricity consumption with comparisons to consumption during a different period, such as the previous day, month, year, etc. Such feedback is received in a number of ways: through detailed electricity bills (see Carroll et al., 2014; Schleich et al., 2013), online via a website (see Benders et al., 2006; Ueno et al., 2006; Gleerup et al., 2010; Vassileva et al., 2012; Mizobuchi and Takeuchi, 2013; Schleich et al., 2013; Harries et al., 2013; Houde et al., 2013), in real-time via a monitor in the home (see Van Dam et al., 2010; Grønhøj and

Thøgersen, 2011; Alahmad et al., 2012; Carroll et al., 2014; Schultz et al., 2015).

The provision of information on individual electricity consumption information allows households to develop a greater awareness of their electricity consumption. By comparing their consumption from one period to another, such information allows households to see which behaviours result in increased consumption, so that they can follow their electricity consuming activities and determine when and how they consume the most electricity, and thus when and how to reduce their consumption.

Hypothesis 2a*: Individual feedback on electricity consumption has a negative effect on electricity demand.*
Hypothesis 2b*: Real-time feedback on electricity consumption has a negative effect on electricity demand.*

A further type of personal feedback that households may receive can be advice tailored to their particular situation (both house and household characteristics) (see Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013) or more general electricity savings tips (see Ueno et al., 2006; Mountain, 2008; Van Dam et al., 2010; Raw and Ross, 2011).

Hypothesis 3a*: Personalised advice on how to save electricity has a negative effect on electricity demand.*
Hypothesis 3b*: Electricity savings tips have a negative effect on electricity demand.*

2.2.2 Social Feedback

Social feedback refers to information on others' electricity consumption, such as neighbours or similar households. It is an intervention which has seen a recent resurgence in popularity in experimental studies and uses the notions of social and injunctive norms. A social norm refers to descriptive consumption feedback of personal consumption compared to that of other households. An injunctive norm reinforces whether a particular behaviour is socially approved or disapproved of. In the case of electricity consumption, an injunctive norm confirms whether a household's consumption is pro-social, i.e. whether the household is a low-consuming household (Schultz et al., 2007).

These two types of social feedback have been separated in the present analysis as there is evidence that solely descriptive comparative feedback leads to a boomerang effect where low-consuming households increase their consumption, converging towards the average (Schultz et al., 2007; Allcott, 2011; Ayres et al., 2012). The inclusion of injunctive norms reinforces the idea that households who consume less than average are engaged in a pro-social behaviour and so they do not increase their consumption.

Hypothesis 4a: *Feedback in the form of social norms does not have an effect on electricity demand.*
Hypothesis 4b: *Feedback in the form of injunctive norms has a negative effect on electricity demand.*

3 Previous Meta-Analyses

The effect of different feedback types and monetary incentives on electricity consumption has been studied by researchers and utilities alike since the 1970s, and as such, several reviews and analyses have been undertaken (see Darby et al., 2006; Ehrhardt-Martinez et al., 2010; Faruqui et al., 2010; Delmas et al., 2013; Faruqui and Sergici, 2013; McKerracher and Torriti, 2013). Table 1 summarises the results of the previous reviews and analyses discussed in this section.

Authors	Objective	Timeframe	Studies	Effect
Darby (2006)	Effect of direct and indirect feedback on energy (gas and electricity) consumption	1979-2006	38	Direct: -15% to -5% Indirect: -10% to 0%
Ehrhardt-Martinez et al. (2010)	Effect of different feedback treatments on energy consumption	1974-2010	57	-12% to -4%
Faruqui et al. (2010)	Effect on IHDs on energy consumption	1989-2010	12	-13% to -3%
Delmas et al. (2013)	Reduction in energy consumption via different treatments	1975-2012	59	-55% to +18% Weighted ATE[1]: -7.4%
Faruqui and Sergici (2013)	Peak demand reduction of time-varying prices.		34	-58% to 0%
McKerracher and Torriti (2013)	Effect of IHDs on energy consumption	1979-2015	27	-5% to -3% ATE: -6.4%

Table 1: Summary of results of previous reviews and meta-analyses

Darby et al. (2006) reviews 38 feedback studies from 1979 to 2006 and concludes that, on average, direct feedback which is received immediately after the energy consuming behaviour is more effective than indirect feedback such as an energy bill. Both Faruqui et al. (2010) and McKerracher and Torriti (2013) analyse the effect of real-time feedback, via an in-home display (IHD), on energy consumption. In a review of 12 pilot studies (1989-2010), Faruqui et al. (2010) find an energy reduction of 18% on average. McKerracher and Torriti (2013) perform a wider analysis of 27 peer and non peer reviewed studies between 1979-2011. The authors find that as sample size increases, the reported treatment effect decreases. Additionally, they classify studies via sampling selection and recruitment method and find that studies with more representative samples report lower percentages of energy reduction.

Hypothesis 5: *As sample sizes increase, smaller effects of incentives on electricity demand are observed.*

Ehrhardt-Martinez et al. (2010) review 57 studies from 1974-2010 covering both feedback and dynamic pricing studies using advanced metering infrastructure. The authors conclude that feedback interventions result in a greater overall reduction in energy consumption than dynamic pricing which is more effective at decreasing demand at peak times.

Focusing on the effect of pricing strategies, Faruqui and Sergici (2013) find that the more dynamic the pricing strategies[2], the greater the amount of peak energy conserved, all the more so when enabling technology is used (Faruqui and Sergici, 2013).

Delmas et al. (2013) provide the most recent analysis of studies from 1975 to 2012 finding that tailored advice and energy conservation tips are most effective at reducing energy consumption. Similarly to McKerracher and Torriti (2013), the authors compare the average treatment effects of studies which use more controls (those which include a control group, demographic information and control for weather changes) to studies with fewer controls studies. They find that higher control studies report a lower reduction in energy consumption (Delmas et al., 2013).

Hypothesis 6: *Higher control studies (inclusion of control group, weather controls, demographic controls, opt-out recruitment, random assignment to*

[2]Real-time pricing strategies are considered to be more dynamic as the price faced by final consumers fluctuates in line with wholesale prices. Time-of-use tariffs are less dynamic as the prices are fixed for certain hours. Critical peak pricing and peak-time rebates fall in-between the two.

treatment group) show a smaller negative effect of an incentive on electricity demand.

Each of these reviews and analyses have covered studies across a long time period, from the seventies and eighties to the present. (Ehrhardt-Martinez et al., 2010, p.42) find trends in energy savings across two distinct periods; the Energy Crisis Era from the seventies to 1995, and the Climate Change Era from 1995 to 2010. (McKerracher and Torriti, 2013, p.393) identify an additional era, from 2005 onwards which they name the Smart Grid Era. The current paper seeks to better understand the effect of different interventions on energy consumption by considering solely studies from 2005 onwards so as to focus on the smart grid era. As (Ehrhardt-Martinez et al., 2010, p.74) note, "studies that compare feedback-related savings across all four decades may result in inflated expectations regarding potential energy savings today".

***Hypothesis 7**: Average effect of incentives on electricity consumption is lower in Smart Grid Era compared to previous eras.*

4 Method

4.1 Data Collection

In order to find appropriate articles for this analysis, the following databases were searched: ScienceDirect, EconLit, Web of Science, SpringerLink, Econpapers, SSRN, NBER, for the following sets of keywords using Boolean logic:

- Keywords concerning type of consumption: electricity consumption, electricity demand, electricity usage, energy consumption, energy demand, energy usage, and;
- Keywords concerning the type of incentive:
  - Incentive, behaviour
  - Informational feedback: smart meter, advanced met*, feedback, nudge, norm,
  - Financial feedback: dynamic pricing, tariff, time of use, critical peak pricing, real time pricing, peak time rebate, and;

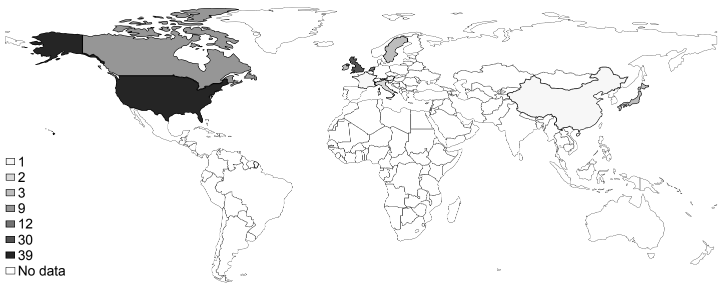

Figure 1: Geographical distribution of included studies

- Keywords concerning the level of consumption: residential, household, consumer, and;
- Keywords concerning the study type: pilot, trial, experiment, field.

Across all databases, after eliminating doubles, the search terms resulted in a list of 1,490 studies. The titles and abstracts of these studies were reviewed. In addition, the reference lists and the lists of citing articles for each selected article, as well as previous meta-analyses, were scanned for further relevant studies. This procedure resulted in a selection of 84 articles and 27 reports on the topic of using incentives to reduce residential electricity consumption. Each article and report was read and a final selection of 24 articles and 15 reports were kept for the analysis.

The final list of articles, those in which the treatment effect is reported as the change in electricity consumption of treated households compared to a baseline or control group, can be found in Appendix A and details on why 72 papers were excluded can be found in Appendix B. A coding protocol was implemented for the final selection of 39 studies which involved an experimentation of the above incentives. The majority of articles came from economics, business, and energy journals. The reports are from utility and government websites as well as from consulting companies.

Figure 1 displays the geographical distribution of included studies. The majority of studies come from the United Kingdom and North America. These countries have be at the forefront of field experiments and pilot studies on incentives to reduce electricity consumption. In addition, this could also be explained by the fact that one of the inclusion criteria is that the paper be written in English and that experiments carried out by national utilities and governments are likely to be written in the native language. This restriction could result in publication bias which will be assessed below.

4.2 Model and Estimation Method

Meta-regression analysis is a quantitative method of systematically analysing the results of empirical studies with a common objective. It goes beyond a literature review in that it allows the analyst to calculate a mean treatment effect across studies by discovering which variables lead to differences in experiments which study the same treatment effect (Stanley and Jarrell, 1989; Nelson and Kennedy, 2009). Meta-analyses are used to estimate a more precise estimate of the true effect of a treatment effect than any single study can do alone (Borenstein et al., 2009).

Using notation from Nelson and Kennedy (2009, p.8), the following meta-regression model is estimated:

$$\tilde{\beta}_i = \alpha_0 + \alpha_1 x_{i1} + ... + \alpha_K x_{iK} + e_i \tag{1}$$

where $(x_{i1}, ..., x_{iK})$ is a vector of study characteristics, $(\alpha_1, ..., \alpha_K)$ are unknown parameters to be estimated, and e_i is the normally distributed sampling-estimation error with zero mean and variance σ_i^2, $\forall i = 1, ..., N$.

This model can be estimated by ordinary least squares (OLS). However, given that in the sample of primary studies, there are treatemnt effects from studies of varied sample sizes, the method of estimation by OLS may lead to inefficient and biased estimates. This bias can be mitgated by using White or Huber-White robust standard errors Sebri (2014).

Furthermore, the standard OLS approach may not be appropriate due to issues highlighted by Nelson and Kennedy (2009) and Stanley and Doucouliagos (2012) which are prevalent in meta regression analysis such as publication bias, heterogeneity, heteroscedasticity and non-independence. Publication bias is an issue across much social science research where results that show a significant effect are favoured for publication over those which do not. Heterogeneity is present due to either differences in the experimental design and methods used in the primary studies, or to differences such as geographical location and historical context. The issue of heteroscedasticity arises from the inclusion of primary studies with different sample sizes, and finally, non-independence occurs when more than one observation is used from a single primary study. Each of the issues are of concern in the present meta-analysis and steps are taken to reduce their impact on the results as discussed below.

Other approaches used in meta-regression analysis to estimate the model in eq. (1) include using fixed- or random-effects estimation (FEE and REE respectively)[3]. FEE weights each treatment effect estimate by its precision squared, or the inverse of its variance. Furthermore, FEE assumes that all primary observations of treatment effects are drawn from the same population (Stanley and Doucouliagos, 2012). In the present sample, treatment effects are taken from primary studies from different countries which thus have different samples. Given such heterogeneity in the sample, Stanley and Doucouliagos (2012) suggest that the REE is a technically more appropriate estimator as the weight used accounts for this heterogeneity.

In further research, Stanley and Doucouliagos (2015) find that the weighted least squares (WLS) estimator is preferable to both FEE and REE. The authors find that under heterogeneity, WLS outperforms FEE, and in the case of publication or small sample bias, WLS does better than REE. Given the characteristics of the data used in the present meta-analysis, several approaches are taken to overcome the potential issues of publication bias, heterogeneity, hetereoscedasticity, and non-independence.

Firstly, to limit issues of publication bias, both peer reviewed articles and reports from the grey literature are included in this analysis. In addition, after a description of the dataset and before any models are estimated, the selection of primary studies used in the meta-analysis is assessed for publication bias. This analysis leads to the conclusion that publication bias is present up to a factor of 2 and that using the sample size as a weight mitigates this problem.

Second, to tackle the sources of heterogeneity, a set of binary variables describing the study characteristics which are potential sources of heterogeneity are included in the regression (section 4.3 describes the variables used in the analysis), the temporal context has been limited to primary studies published since 2005 representing the *Smart-Grid Era* (McKerracher and Torriti, 2013), and additional data regarding the location of the primary studies has been collected.

Next, to account for heteroscedasticity, the model in eq. (1) is estimated by WLS. The preferred weight is the inverse standard error of the treatment effect, however, given that these are not always reported in the primary studies, a common approach is to proxy the standard error using the sample size (Nelson and Kennedy, 2009; Stanley and Doucouliagos, 2012). As such, the square root of the sample size is used as weights for the estimation following (Delmas et al., 2013; Sebri, 2014; Van Houtven et al., 2017) such that experiments with a larger sample are given more weight. Experiments with larger samples are considered to be more representative of the population and so

[3]These terms refer to estimators used in meta-analysis and not to those used in panel data econometrics (Stanley and Doucouliagos, 2012).

the estimated effect is a better estimate of the true effect.

Finally, to address the non-independence of several treatment effects coming from the same primary study, the estimated standard errors are clustered by primary study.

4.3 Variables

4.3.1 Dependent Variable

The variable of interest is the treatment effect reported in primary studies as the percentage change in electricity consumption as a result of the implementation of an incentive. When a control group is present in an experiment, the percentage change relative to the control group is used. If no control group is present, the percentage change relative to the baseline is used[4]. A negative percentage indicates a reduction in electricity consumption, whereas a positive percentage change indicates an increase in electricity consumption.

4.3.2 Independent Variables

The independent variables refer to the type of intervention tested in the primary study and the controls used. As discussed above, there are *monetary incentives*: households receive a financial reward which is directly linked to their electricity conservation effort. For example, changing prices are used to influence consumers electricity consumption by aligning the retail price of electricity with the wholesale price. Or participating households are given feedback on how much their electricity consumption costs (*monetary information*).

Non-monetary strategies are separated into those which provide personal feedback, and those which provide social feedback of others' electricity consumption. *Individual feedback* refers to interventions where participants receive information on their current and previous consumption in energy units. This refers to consumption information that is in addition to the standard electricity bill, be it a more detailled bill, or consumption information on a web site. *Real-time feedback* refers to the same type of information which is delivered in real-time via an energy monitor.[5] Households can also receive

[4]Presence of a control group is controlled for in the analysis to come.

[5]Only data that are received via an IHD or monitor are considered to be *real-time feedback* in the present analysis. Real-time data are made available to households via

personalised advice specific to their living situation on how to lower their electricity consumption, or generic electricity *savings tips*.

Studies which provide social feedback are separated into those which provide *social norms* feedback: descriptive feedback of personal consumption compared to that of other households, and *injuctive norms* feedback which provides social approval or disapproval of a household's consumption behaviour.

Finally, a set of control variables are included in the analysis: *control group*: presence of a control group; *weather controls*: whether weather is controlled for; *demographic controls*: the collection demographic information; *random*: households are assigned randomly to control and treatment groups as opposed to choosing an intervention; *opt-in recruitment*: households choose to participate in the study; and *duration*: duration of study. These control variables are included in order to capture the heterogeneity between the different experiments. Furthermore, studies which include such controls can be judged to be of higher quality as they control for changes in behaviour which cannot be explained by the use of an incentive alone.

5 Results

5.1 Descriptive Statistics

The analysis covers 105 observations from 39 unique papers giving, on average, 2.7 observations per paper. In meta-analysis it is preferable to limit the analysis to one observation per study in order to reduce correlation between studies (Nelson and Kennedy, 2009). However, given that some reports describe the results of more than one experiment, and also, due to the design of the sample experiments, doing so would greatly limit the number of observable treatment effects. To account for potential heterogeneity due to several observations being taken from one study, in the following analysis, standard errors are clustered by study.

Table 2 provides descriptive statistics of the independent and dependent variables for the full sample. Within the sample of studies selected for this analysis, *individual feedback* is the most experimented treatment. Compared with previous meta-analysis, the share of studies involving a form of social feedback (*social norms* or *injunctive norms*) has increased. The *indi-*

websites (see Houde et al., 2013), however, the data are not accessible to consumers in real-time. They must log-on to the site in order to access the information. The incentives used in such experiments are included in *individual feedback*.

vidual feedback treatment represents 70% of the observations and 77% of the studies, and the *injunctive norms* treatment represents 27% and 26% of the observations and studies, respectively.

Concerning the design of the primary studies, the majority use a control group for comparison and control for demographic differences in the sample population, 90% and 85% respectively. Fewer studies (59%) control for variations in the weather. 68% of observations randomly assign subjects to a treatment but this is not a practice adopted in all studies, 49%. Opt-in recruitment is the more common method of recruitment, 67% of observations and 69% of studies.

5.2 Average Effects by Treatment

Table 2 also provides both a non-weighted and weighted average treatment effect (ATE) by incentive. The treatment effects are weighted using study sample size as frequency weights following Schmidt and Hunter (2014) which gives more weight to studies with larger samples. The ATE across all incentives is -3.37%. The weighted ATE takes into consideration the differing sample sizes in each study and equates to a 1.85% reduction in electricity consumption. This means that, on average, an incentive in a typical electricity conservation study will result in electricity savings of slightly less than 2%. In the sample of studies selected, the effect of incentives on electricity consumption ranges from an 22.2% reduction (Kendel and Lazaric, 2015) to a 13.69% increase (Torriti, 2012).

From table 2, it can be seen that *real-time feedback* and *monetary information* have the greatest effects on electricity consumption with a weighted average reduction in consumption of 2.89% and 2.86%, respectively. *Monetary incentives* have the smallest effect on electricity consumption with a weighted average reduction in consumption of 0.99%.

Study characteristic	Mean	Std. dev.	Primary obs.	Primary studies	Average sample size	Min (%)	Max (%)	ATE (%)	Weighted ATE (%)
treatment effect					6685	-22.20	13.69	-3.37	-1.85
Monetary incentive	0.34	0.48	34%	28%	1986	-7.60	13.69	-2.57	-0.99
Monetary information	0.31	0.47	31%	46%	716	-18.06	5.30	-4.18	-2.86
Individual feedback	0.70	0.46	70%	77%	9429	-22.20	5.30	-3.56	-1.88
Real-time feedback	0.37	0.49	37%	38%	566	-18.06	5.30	-4.69	-2.89
Personalised feedback	0.18	0.39	18%	13%	19504	-5.80	-1.20	-2.22	-2.01
Savings tips	0.52	0.50	52%	46%	4706	-16.71	5.30	-3.30	-1.78
Social norms	0.10	0.29	10%	21%	530	-18.00	-0.35	-4.66	-1.74
Injunctive norms	0.27	0.44	27%	26%	21720	-7.02	-1.00	-2.26	-1.95
Control group	0.90	0.31	90%	85%	7445				
Weather controls	0.73	0.44	73%	59%	6128				
Demographic controls	0.85	0.36	85%	79%	6511				
Random assignment	0.68	0.47	68%	49%	9218				
Opt-in receruitment	0.67	0.47	67%	69%	540				
Duration (months)	13.60	8.81	100%	100%	6685				
Sample size	6685	14863	100%	100%	6685				
Number of observations			105	39					

Table 2: Descriptive statistics and average treatment effects

Incentive	Weighted ATE Peer reviewed (%)	Grey literature (%)
Overall	-1.96	-1.71
Monetary incentive	2.31	-1.25
Monetary information	-3.63	-2.77
Individual feedback	-2.02	-1.72
Real-time feedback	-2.83	-2.89
Personalised advice	-2.01	
Savings tips	-3.01	-1.76
Social norms	-2.36	-1.12
Injunctive norms	-2.01	-1.85
Number of observations	57	48

Table 3: Comparison of weighted average treatment effects

For comparison between the literature types, table 3 provides the weighted average treatment effects by study type, i.e.: whether the study is from a peer-reviewed journal or from the grey literature. Across all incentive types, on average, a peer-reviewed study shows a weighted ATE of -1.96%, and a study from the grey literature shows a weighted ATE of -1.71%. In the sample of studies collected, there are no reports which use personalised feedback as an incentive. Grey literature studies tend to show a smaller effect of an incentive on electricity consumption. Among the peer reviewed sutdies, the weighted ATE of the use of *monetary incentives* is an increase in electricity consumption of 2.31%.

The primary studies are separated into those which use a higher number of controls; a control group, weather and demographic controls, randomly assign households to treatments, and use an opt-out method of recruitment, as such studies are assumed to show a more representative estimate of the true treatment effect. Studies which compare the treatment effect to a control group rather than the baseline of the same group of households, provided a more robust estimate of the treatment effect. The same applies to studies which use weather controls and collect demographic information.

Studies which adopt a random treatment assignment method and an opt-out method of recruitment are more representative as they use samples which have not chosen their treatment method nor are subject to selection bias.

Table 4 gives the average treatment effects by study control level. High control studies are considered to be those which include all the above controls, lower control studies are those which include less. Of all the studies, 22% can

	Primary obs.	ATE (%)	Min (%)	Max (%)
All studies	105	-3.37	-22.20	13.69
Higher control studies	23	-2.17	-5.40	-1.17
Lower control studies	82	-3.71	-22.20	13.69

Table 4: Average treatment effects by study quality

be considered to use a high level of controls. The high control studies have an ATE of -2.17% whereas the lower control studies have an ATE of -3.71%. A test of whether the ATE are equal can be rejected at the 1% level.

Table 5 provides the correlations between variables. There are no strong correlations between *treatment effect* and the treatment variables. Strong positive correlations can be seen between both the *personalised feedback* and the *social norm* and *injunctive norm* treaments, and strong negative correlation with *opt-in recruitment* as for these treatments, participating households took part in the study by default and opted-out if they did not want to take part. These studies are typically large-scale experiments led by utilitieswhich have the means to carry out such studies (Allcott, 2011; Ayres et al., 2012).

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) treatment effect	1.00																
(2) Monetary incentive	0.13	1.00															
	(0.17)																
(3) Monetary information	-0.13	0.07	1.00														
	(0.20)	(0.46)															
(4) Individual feedback	-0.07	-0.31*	0.27*	1.00													
	(0.50)	(0.00)	(0.01)														
(5) Real-time feedback	-0.23*	0.03	0.63*	-0.09	1.00												
	(0.02)	(0.79)	(0.00)	(0.36)													
(6) Personalised advice	0.13	-0.34*	-0.32*	0.26*	-0.36*	1.00											
	(0.20)	(0.00)	(0.00)	(0.01)	(0.00)												
(7) Savings tips	0.02	0.09	0.07	-0.09	0.26*	-0.49*	1.00										
	(0.87)	(0.38)	(0.48)	(0.35)	(0.01)	(0.00)											
(8) Social norms	-0.10	0.04	0.06	0.00	-0.12	-0.15	0.11	1.00									
	(0.32)	(0.69)	(0.54)	(0.97)	(0.24)	(0.12)	(0.24)										
(9) Injunctive norms	0.15	-0.44*	-0.41*	0.35*	-0.42*	0.67*	-0.20*	-0.20*	1.00								
	(0.11)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.05)									
(10) Control group	0.27*	0.18	0.10	0.25*	-0.12	0.16	-0.14	-0.10	0.21*	1.00							
	(0.01)	(0.06)	(0.32)	(0.01)	(0.21)	(0.10)	(0.16)	(0.31)	(0.03)								
(11) Weather controls	0.17	0.16	0.08	-0.02	-0.07	0.17	-0.19	-0.32*	0.22*	0.36*	1.00						
	(0.09)	(0.10)	(0.40)	(0.80)	(0.47)	(0.08)	(0.06)	(0.00)	(0.03)	(0.00)							
(12) Demographic controls	-0.13	-0.25*	0.06	0.18	-0.00	0.20*	0.02	0.14	0.14	-0.06	0.10	1.00					
	(0.20)	(0.01)	(0.55)	(0.07)	(0.97)	(0.04)	(0.84)	(0.16)	(0.17)	(0.55)	(0.29)						
(13) Random assignment	0.17	-0.01	-0.10	0.29*	-0.27*	0.27*	-0.05	-0.26*	0.37*	0.49*	0.41*	0.39*	1.00				
	(0.09)	(0.88)	(0.30)	(0.00)	(0.01)	(0.00)	(0.62)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)					
(14) Opt-in recruitment	-0.22*	0.30*	0.39*	-0.29*	0.38*	-0.56*	0.13	0.09	-0.81*	-0.18	-0.11	0.15	-0.19	1.00			
	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.17)	(0.35)	(0.00)	(0.07)	(0.28)	(0.13)	(0.06)				
(15) Duration	0.19	-0.18	0.04	0.22*	-0.07	0.31*	-0.08	-0.05	0.24*	0.11	0.31*	0.28*	0.43*	-0.13	1.00		
	(0.05)	(0.06)	(0.70)	(0.02)	(0.47)	(0.00)	(0.41)	(0.63)	(0.01)	(0.27)	(0.00)	(0.00)	(0.00)	(0.19)			
(16) Sample size	0.16	-0.23*	-0.27*	0.28*	-0.32*	0.41*	-0.14	-0.14	0.61*	0.15	-0.06	-0.03	0.25*	-0.59*	0.20*	1.00	
	(0.10)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.15)	(0.17)	(0.00)	(0.13)	(0.53)	(0.78)	(0.01)	(0.00)	(0.04)		
(17) Peer reviewed	0.02	-0.30*	-0.33*	-0.07	-0.28*	0.43*	-0.26*	0.10	0.25*	-0.31*	-0.16	0.14	-0.14	-0.08	-0.26*	-0.00	1.00
	(0.87)	(0.00)	(0.00)	(0.49)	(0.00)	(0.00)	(0.01)	(0.30)	(0.01)	(0.00)	(0.09)	(0.15)	(0.14)	(0.41)	(0.01)	(0.97)	

Standard errors in parentheses. $^{*}p < 0.05$

Table 5: Pearson cross-correlation table

Figure 2 shows the distribution of treatment effect by publication year. The majority of studies were published from 2010 onwards. Almost half of the observations in the sample were published in 2011. There does not appear to be a trend in the effects of incentives on electricity consumption over this time period.

Figure 2: treatment effect by year of publication

Figures 3 to 5 show how treatment effect varies in relation to the use of a control group, the control of weather effects, or the collection of socio-demographic variables in the primary studies. These figures show that, bar one high, positive observation, the distribution of treatment effects is similar whether a control group is present, whether weather effects have been controlled for, and whether socio-demographic data is collected or not.

Figures 6 to 8 are box plots of the spread of treatment effects by the presence of a control group, the use of weather controls, or the collection of socio-demographic data. Figure 6 shows that the median treatment effect is slightly smaller when a control group is present, and that the spread is greater in the absence of a control group. Whether weather effects are controlled for or not, the median treatment effect is similar. The spread is slightly tighter around the mean when weather is controlled for. Concerning the collection, or not, of socio-demographic data, the median and the spread of the treatment effects are similar. From these box plots, there is evidence of certain outlying values of the treatment effects.

Figures 9 and 10 provide scatter plots of the distribution of treatment effect by treatment assignment method (random assignment) and recruitment

Figure 3: Treatment effect by presence of control group

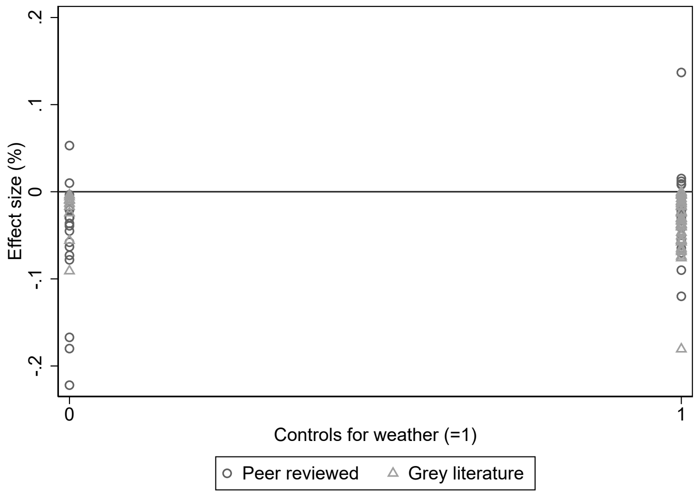

Figure 4: Treatment effect by weather controls

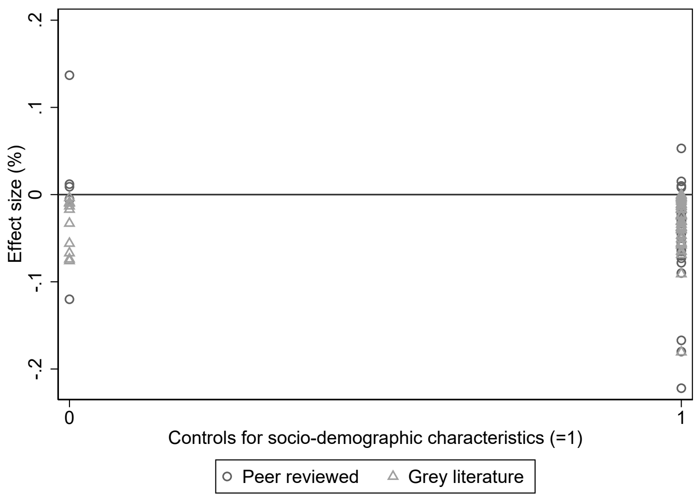

Figure 5: Treatment effect by collection of socio-demographic data

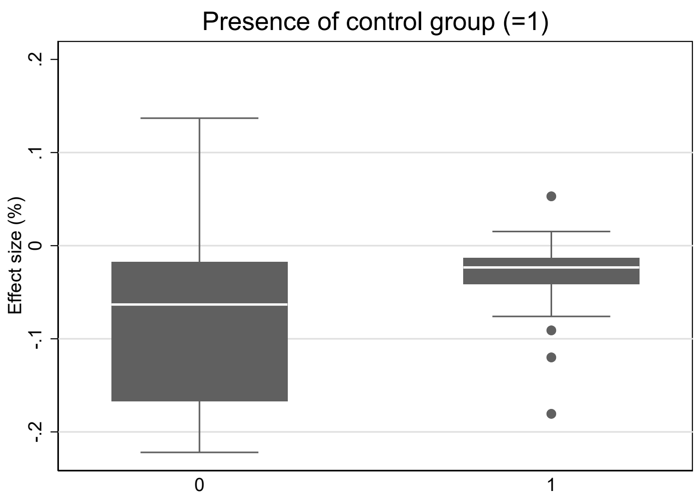

Figure 6: Box plots of treatment effect by presence of control group

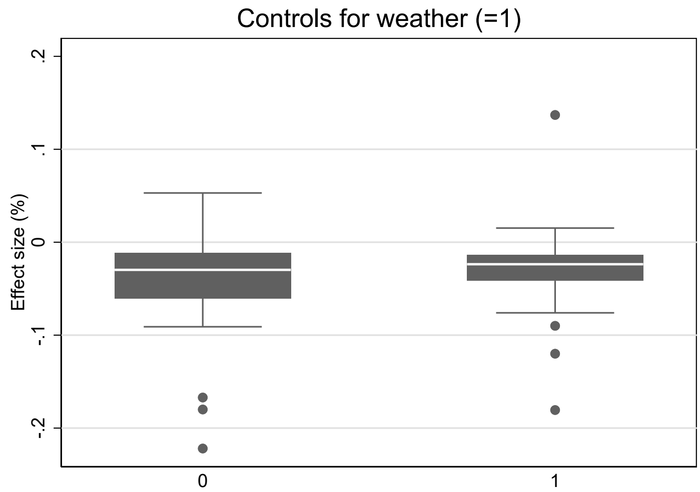

Figure 7: Box plots of treatment effect by use of weather controls

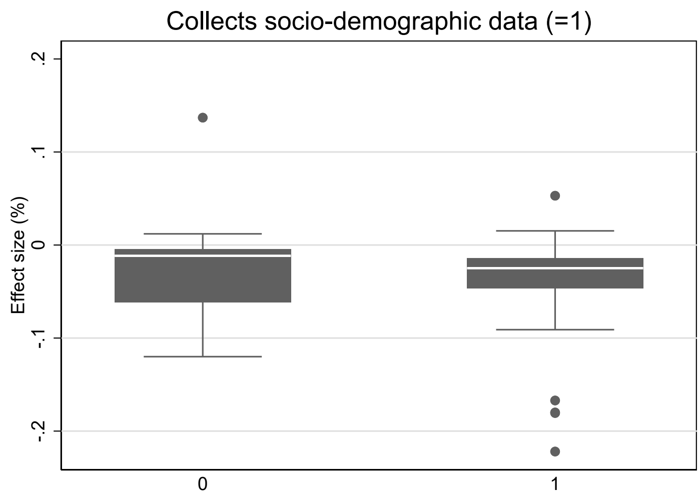

Figure 8: Box plots of treatment effect by collection of socio-demographic data

method (opt-in) for both literature types. For both literature types, approximately two-thirds of the sample studies use random assignment and/or opt-in methods. The distribution of treatment effect by treatment assignment method shows a greater dispersion in electricity consumption change when subjects are not randomly assigned to a treatment, with the exception of one observation. In addition, we can see a greater reduction in electricity consumption when subjects self-select into a treatment (Figure 10) than when they do not.

Figures 11 and 12 are box plots showing the spread of the data by treatment assignment method and by sample selection method. In both cases, the median values are similar, however, the spread is more closely concentrated around the median values when treatment assignment is random and when participants must opt-out of the study. Households can achieve greater levels of electricity consumption reduction when they are not randomly assigned to a treatment and when they choose to participate in a study.

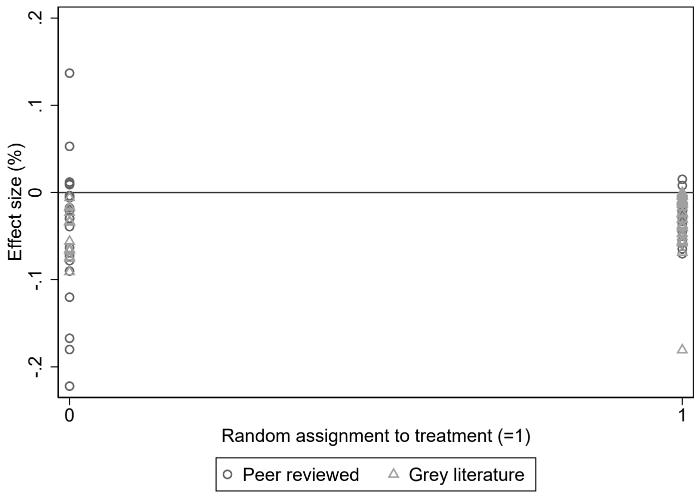

Figure 9: Treatment effect by treatment assignment method

Figure 13 shows the distribution of treatment effect by duration of the study. The majority of studies are short in duration (shorter than 12 months). There are a cluster of studies lasting one or two years. The majority of the longer studies are those that are led by utilities. Finally, there are few utility led studies which last for almost three years. From the figure, it appears that longer studies show a smaller effect of incentives on electricity savings.

Finally, fig. 14 shows treatment effect against literature type. Whilst the greatest reduction for each literature type is similar, the range of treatment

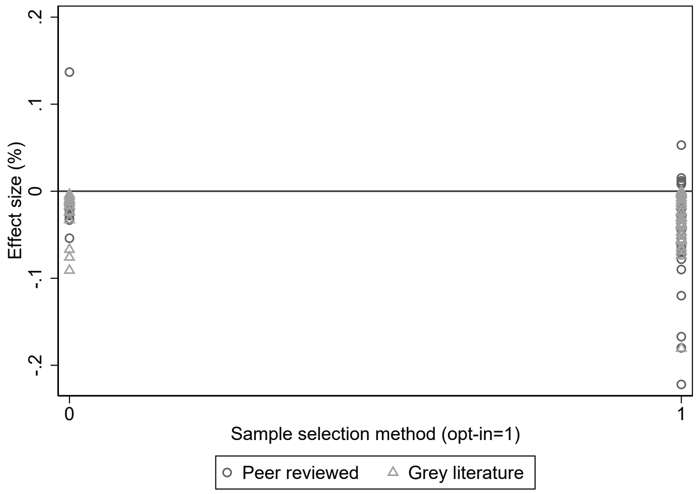

Figure 10: Treatment effect by sample selection method

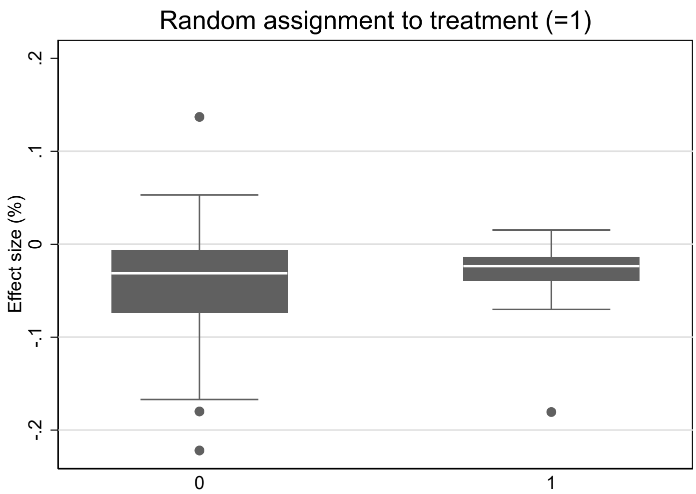

Figure 11: Box plots of treatment effect by treatment assignment

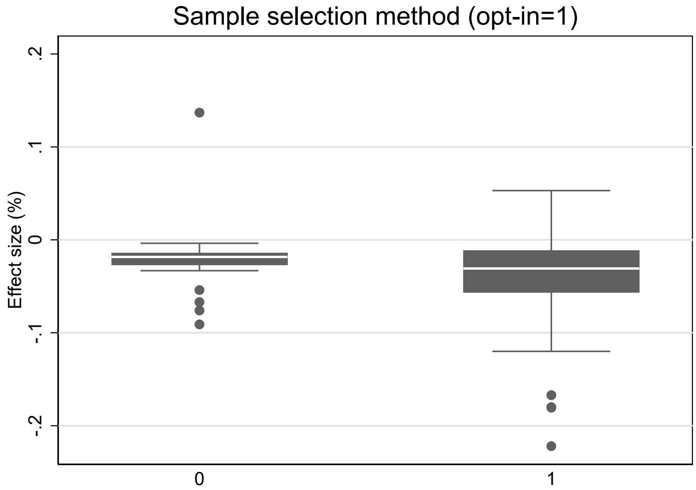

Figure 12: Box plots of treatment effect by sample selection method

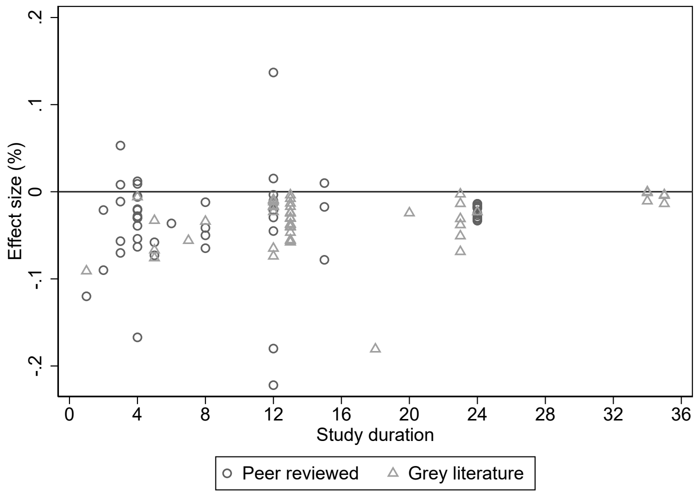

Figure 13: treatment effect by study duration

effects is greater for peer reviewed literature and there are more positive results (increase in electricity consumption) for peer reviewed literature.

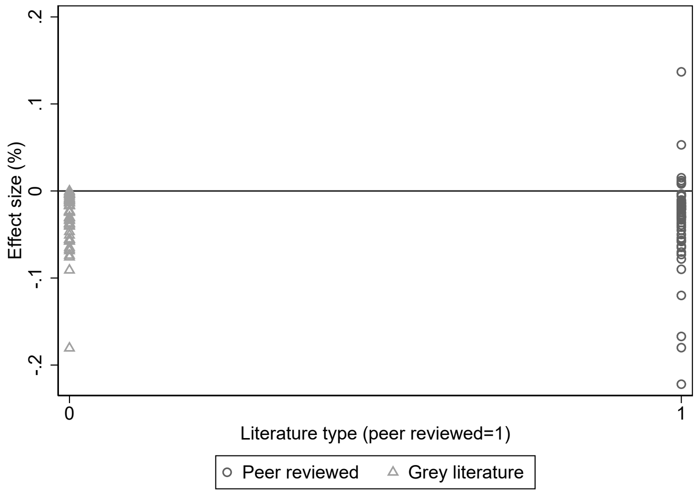

Figure 14: treatment effect by literature type

The above graphical analysis indicates that the treatment effects reported in primary studies may be particularly affected by the presence of a control group, treatment assignment and sample selection methods.

In studies without a control group, the change in electricity consumption is compared within the same group of households between the treatment period and a baseline period. Whereas in studies with a control group, the change in consumption is compared both within the same group of households and between groups of households whose consumption is measured during the treatment and baseline periods; a difference-in-difference method. The latter studies allow researchers to account for additional factors which affect electricity consumption during the course of the duration of the study and appear to show a lesser treatment effect to the former.

Households who choose to participate in a study on electricity consumption may be particularly motivated to reduce their consumption. Those who participate in studies on an opt-out basis (which is arguably more representative of a national roll-out of such interventions) achieve much smaller levels of electricity reduction.

When households are randomly assigned to treatment groups, they achieve smaller electricity savings than when they are not. This would suggest that a tailored approach to treatment design corresponding to households existing motivations to change their electricity consumption is pertinent. Such motivations maybe monetary, or for environmental reasons, or other.

The inclusion of weather controls and the collection of socio-demographic data does not appear to have a strong impact on the reported treatment effects.

The impact of these study design choices on the treatment effects will be further analysed in section 5.4.

5.3 Publication Bias Analysis

In this section we explore the sample of primary studies used in this meta-analysis for issues of publication bias. According to Card and Krueger (1995) there are three potential sources of publication bias in economic research: (1) a predisposition to accept studies which are consistent with the conventional view; (2) reported models may be selected based on the presence of a conventionally expected results; (3) a tendency to publish only statistically significant results.

Potential publication bias in the sample of primary studies used in this meta-analysis can be analysed graphically using a funnel plot, as shown in fig. 15. These graphs plot treatment effects against a measure of precision, such as the inverse standard error of the treatment effect or the sample size of the treatment group. The intuition is that the accuracy of the treatment effect increases with the level of precision. Studies with larger standard errors and smaller sample sizes are dispersed at the bottom of the graph, with the spread of treatment effects decreasing as standard errors decrease and sample sizes increase. In the absence of publication bias, the result is a symmetrical, inverted funnel shaped graph. On the other hand, if there is a publication bias, an asymmetrical funnel can result due to an absence of publications of non statistically significant results (Egger et al., 1997; Sterne et al., 2004).

The funnel plot in fig. 15 plots treatment effect against the square root of sample size. The plot shows that the majority of treatments result in a negative effect on electricity consumption. No studies from the grey literature report an increase in electricity consumption and there are more observations from peer-reviewed articles dispersed at the bottom of the funnel. The somewhat asymmetrical nature of the funnel plot suggests that there may be an issue of publication bias in the present analysis due to some results not being included in the analysis.

Stanley et al. (2010) suggest that publication bias may be reduced and scientific inference improved by averaging the treatment effects of the top 10% of the funnel as these are the most precise estimates. Table 6 shows the average and weighted average treatment effects for the full sample and the

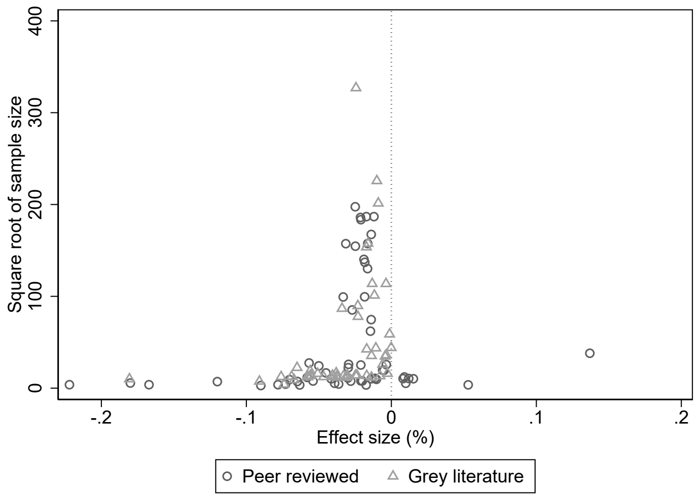

Figure 15: Funnel plot of treatment effect versus sample size

top decile according to the weight used[6]. Comparing the average treatment effects for the top 10% of the funnel and the full sample suggests that, on average, the effect of incentives on electricity consumption is overestimated by a factor of 2. When sample size is accounted for, as the weighted ATE shows, the distortion due to publication bias is greatly reduced and the difference is not significant (p=0.8641).

As the inverse standard error is the preferred measure of precision, we also report the average treatment effects of the 42 observations for which standard errors are reported or can be constructed. The distortion due to publication bias is smaller for this subset of the sample when comparing ATE (a factor of 1.8), and the difference in values in not significant (p=0.8022) once sample size is accounted for.

	ATE (%)		Weighted ATE (%)	
	Sample size	1/SE	Sample size	1/SE
Top 10% of funnel plot	-1.69	-1.69	-1.79	-1.62
Full sample	-3.37	-3.06	-1.85	-1.75

Table 6: ATE correcting for publication bias

The above correction for publication bias suggests that if present, any

[6]Where the inverse standard error is used as a weight, there are only 42 observations in the sample as the standard error is not available for all studies. We use this sub-sample as a robustness check for issues of publication bias as standard error is the preferred weight.

bias is small and not statistically significant once sample sizes have been accounted for in calculating weighted average treatment effects. Nevertheless, it is prudent to test for the existence of such bias.

In the presence of publication bias, treatment effects are positively correlated with their standard errors (Stanley and Doucouliagos, 2012). This suggests that the size of an effect will depend on its standard error:

$$treatment\_effect_i = \beta_0 + \beta_1 SE_i + \epsilon_i \tag{2}$$

To account for differences in the primary studies, the equation is weighted by a measure of precision, ideally the inverse of its standard error (Stanley et al., 2010):

$$t_i = \beta_0(1/SE_i) + \beta_1 + v_i \tag{3}$$

where t_i is the t-statistic of the treatment effect. As standard errors are not available for all observations, we also reconstruct this equation using the square root of sample size as the measure of precision:

$$treatment\_effect_i/sample\_size_i^{0.5} = \beta_0(1/sample\_size_i)^{0.5} + \beta_1 + v_i. \tag{4}$$

In the presence of publication bias, treatment effects are positively correlated with their standard errors, and negatively correlated with sample sizes, as standard errors are inverse functions of sample size (Stanley and Doucouliagos, 2012; Schmidt and Hunter, 2014). Estimates of β_0 from eqs. (3) and (4) are an alternative correction of publication bias (Stanley and Doucouliagos, 2012). Table 7 shows the results of the model in eqs. (3) and (4) as well as for the full sample using the square root of sample size as a proxy measure of precision.

Testing $H_0 : \beta_1 = 0$ is a test of whether publication bias is present, the funnel asymmetry test. In each of the three estimations, the null hypothesis cannot be rejected, there is no evidence of publication bias for the subset using the two different weights, nor for the full sample.

A second test, the precision effect test, of whether there is a genuine empirical effect can be tested: $H_0 : \beta_0 = 0$. In both models, the null hypothesis is rejected, implying that there is a genuine empirical effect which merits further analysis.

Graphically, the funnel plot suggests that there is a potential publication bias. When comparing the ATE of the full sample to the top 10% of the funnel, this bias is of a factor 2. However, accounting for sample sizes reduces the bias to an small and statistically insignificant amount. Furthermore, tests on both a subset of the sample (for which standard errors are available) and on the full sample, lead to the conclusion that publication bias is not an issue once sample size has been accounted for. Nevertheless, as discussed above, a WLS estimation will e used to account for heteroscedasticity in the sample of primary observations.

	(1) Standard error *Equation (2)*	(2) Sample size *Equation (2)*	(3) Sample size *Equation (3)*
β_0	-1.578***	-7.040***	-7.752***
	(0.305)	(0.204)	(1.909)
β_1	-32.499	0.015*	0.015*
	(40.713)	(0.008)	(0.009)
Observations	42	42	105
R^2	0.777	0.577	0.501

Standard errors in parentheses

Standard errors are clustered by primary study.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimation of publication bias

5.4 Effects of Individual Incentives

The analysis of publication bias has shown such bias to be less of an issue once sample sizes are taken into consideration. We therefore use the square root of sample sizes as weights in the following section in which the effects of the different incentives on electricity consumption are analysed[7].

Table 8 shows the results of the WLS meta-regression analysis across the different incentive types. Specifications 1-3 focus on a particular incentive strategy (monetary, personal feedback and social feedback). The fourth considers the study design features and the final specification includes all variables. Each estimation includes a variable accounting for the duration of the study and the type of literature it is from. Finally, standard errors for each estimation are clustered by study to account for any dependence between studies. Coefficients on the different incentives are interpreted as a

[7]The results of a cluster-robust OLS estimation are provided in Appendix C as a benchmark for the following WLS estimaion.

change in electricity consumption relative to the consumption of the control group, when present in the study which is the case for 90% of the observations, or the baseline level of consumption. A negative coefficient signifies a reduction in electricity consumption.

Monetary incentives have a significant positive effect of reducing electricity savings by 2.8 percentage points. When all incentives are controlled for, this significant effect falls out. The effect of *monetary information* becomes significant, showing a reduction in electricity savings of 2.5 percentage points. These results are opposite to those predicted by the theory. It may be that as *monetary incentives* such as dynamic pricing provide households with the possibility of consuming at a lower price during off-peak periods[8], the rebound effect of consumption outweighs the savings encouraged by the higher peak price (Geelen et al., 2013; Khan et al., 2016). An explanation as to why *monetary information* does not have the predicted effect is that the possible savings are too small to be motivating (Hargreaves et al., 2010; Goulden et al., 2014), or that households expenditure on electricity is small relative to their income (Faruqui et al., 2010; Schleich et al., 2013).

In both the personal feedback and the full specification *individual feedback* has a significant negative effect of 3-4 percentage points on electricity consumption. When such feedback is delivered in real-time no additional significant effects on electricity consumption are found. This suggests that personal feedback on a household's past consumption does not necessarily need to be available in real-time to motivate electricity saving behaviour. *Savings tips* show a significant positive effect on electricity consumption of 4 percentage points. Generic advice on how to save electricity appears to not be effective at reducing consumption. One reason for this is that householders generally know what they should do to reduce their consumption and that reminding them of such behaviours serves to crowd out any intrinsic motivation they had to do so.

In this meta-analysis, comparative feedback is separated into descriptive feedback and feedback which integrates social norms. In specification 3, a significant negative effect of both types of feedback on electricity consumption of 4-5 percentage points is found. This provides new evidence of the effectiveness of such feedback compared to findings in Delmas et al. (2013) who found no significant effect of such feedback. Since their meta-analysis, there has been an increase in large-scale studies of such incentives.

Across the five specifications, the 10% of studies which do not use a con-

[8]Studies which used such incentives were included in the present meta-analysis as the primary authors also considered the effect of the incentive on overall household electricity demand.

trol group show a greater increase in electricity reduction of between 7.3 and 11.2 percentage points compared to those that do use a control group. This suggests that when electricity savings are calculated compared to a baseline of the same group, they may be overestimated. *Duration* of the study has a small significant positive effect on electricity consumption in specifications 1-3, and 5. This adds to the previous evidence that electricity conservation experiments are subject to attrition of the effects of incentives over time Delmas et al. (2013). The positive coefficient on *peer reviewed* suggests that peer reviewed experiments are more conservative in their estimations of the effects of incentive on electricit consumption than those from the grey literature.

6 Discussion

The meta-analysis presented in this paper provides a comparison of different incentives used in the experimental literature to incentivise residential consumers to lower their electricity demand. Contrary to previous analyses, it provides a comparison of contemporary experimental studies by focusing on studies from 2005 onwards, the "Smart Grid Era". Previous analyses risk overstating the potential of different incentives by including older studies McKerracher and Torriti (2013). By restricting the time frame, the intention is to limit the analysis to studies with similar available energy monitoring technology, in order to avoid exacerbating issues of heterogeneity due to differing temporal contexts. In order to avoid issues of publication bias, the present meta-analysis adopted a wide search method to collect data from both peer-reviewed and grey literature studies. To verify the extent of the publication bias issue in the sample of studies used, a detailed analysis of the potential bias was carried out. Indeed, a graphical examination of the potential publication bias suggested that this may be an issue, however estimations of the amount of bias and tests of its presence have shown it to not be a significant issue for the present sample of studies once sample size is considered. Furthermore, the precision effect test shows that there is a genuine underlying effect of interest.

In addition, the experimentation of new methods of encouraging households to lower their electricity demand are included in the present meta-analysis, namely the use of injunctive norms in addition to social norms. Furthermore, a greater level of study design controls are included as controls for heterogeneity between studies. This provides an opportunity to disentangle the effects of such incentives and a more extensive comparison of the effects of different study methods on residential electricity demand.

In this section we discuss the above results and examine whether we can

	(1) Monetary	(2) Personal feedback	(3) Social feedback	(4) Study design	(5) All incentives
Monetary incentive	2.790*				1.318
	(1.462)				(1.571)
Monetary information	0.662				2.492*
	(1.384)				(1.414)
Individual feedback		-3.115**			-3.919**
		(1.358)			(1.675)
Real-time feedback		-0.651			-2.138
		(1.415)			(1.584)
Savings tips		4.385**			4.069**
		(2.104)			(1.967)
Personalised advice		0.562			-0.746
		(2.021)			(2.425)
Social norms			-4.316*		-4.518**
			(2.387)		(2.174)
Injunctive norms			-5.000**		-3.238
			(1.998)		(3.281)
Control group	7.278**	10.790***	8.483**	7.642**	11.161***
	(3.307)	(3.259)	(3.414)	(3.489)	(2.840)
Weather controls	-0.095	0.804	0.856	0.671	-0.671
	(1.436)	(1.449)	(1.385)	(1.311)	(1.985)
Demographic controls	1.295	1.314	2.524	1.104	2.455
	(2.631)	(2.857)	(3.118)	(2.962)	(2.776)
Random assignment	-1.704	-2.727	-1.642	-1.490	-2.783
	(2.216)	(2.446)	(2.777)	(2.419)	(2.457)
Opt-in recruitment	-1.604	0.546	-3.795*	-0.466	-3.262
	(1.554)	(1.710)	(2.179)	(1.336)	(2.840)
Duration	0.198*	0.265**	0.205*	0.170	0.325***
	(0.102)	(0.119)	(0.105)	(0.103)	(0.111)
Peer reviewed	4.638***	4.208**	4.831***	3.503**	5.801***
	(1.698)	(1.840)	(1.635)	(1.549)	(1.883)
Constant	-15.394***	-19.722***	-14.319***	-14.936***	-17.496***
	(4.160)	(5.760)	(4.352)	(4.201)	(5.348)
Observations	105	105	105	105	105
Adjusted R^2	0.195	0.342	0.218	0.181	0.381

Standard errors in parentheses
Inverse square roots of sample size are used as analytical weights.
Standard errors are clustered by primary study.
A negative coefficient reads as a reduction in energy consumption.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: WLS estimation of treatment effects

confirm or reject the hypotheses presented in section 2.

The analysis has shown that on average and before taking into consideration primary study sample size, the different incentives show a negative effect on electricity consumption ranging from -4.69 to -2.22%. Across all incentives, a study on the effect of an incentive on electricity consumption can expected to show a 3.37% reduction in electricity consumption. This effect is lower than that reported in previous meta-analyses, however it is in line with the conclusion of McKerracher and Torriti (2013) that there is a downward trend in the size of conservation effects. Once sample sizes are accounted for, the weighted average treatment effect is 1.85%. We have seen from the analysis of publication bias that accounting for primary study sample sizes goes a lot of the way towards correcting for the bias.

In terms of the level of controls used in the studies, compared to previous meta-analyses, there has been an increase in the number of studies using control groups, and controlling for demographic variables and weather variations which leads to lower, but arguably more reliable, estimates of the effect of incentives on electricity consumption. The two differing levels of study controls show a difference in estimated electricity savings of 1.54 percentage points with studies with more controls (those including control groups, controlling for weather changes and socio-demographic variables, assigning households randomly to treatment groups and using an opt-out method of recruitment), having a smaller average treatment effect, -2.17% versus 3.71%. If such incentives are to be implemented at a national level, these higher control studies may be a better reflection of the level of electricity savings that may be achieved.

A graphical analysis showed that in studies in which households are randomly assigned to a treatment there is a smaller spread in treatment effects. This provides further support for the idea that a national roll-out of a particular incentive may not be the best approach as greater electricity savings can be attained if households are able to pick an incentive which is appropriate to them. A one-size-fits-all may not be the most effective. However, a tailored approach may not be feasible. More research needs to be done in this area to determine whether households are able to pick appropriate incentives, and on the effect of tailored incentives on electricity consumption.

In studies in which households choose to participate, there is a greater spread of treatment effects. These households may have motivations to take part in electricity consumption field experiments and pilot studies that are not necessarily accounted for in the experiment. These participants may be predisposed to make a greater effort than if the incentive were to be implemented at a national level Alexander (2010); Ericson (2011). This implies that caution should be exercised when viewing the results of experiments in which participants self-select into a treatment.

While at the descriptive level, all incentives result in a reduction of residential electricity consumption on average, the econometric analysis shows that only certain incentives have a significant effect once other variables are controlled for. *Monetary*-based incentives have a small positive effect on residential electricity consumption. Hypotheses 1a and 1b can be rejected for the present sample.

Individual feedback has a significant effect at reducing electricity consumption, however, there is no significant effect of *real-time feedback*, when other informational feedback and study design variables are controlled for. Concerning the two types of guidance that can be given to households, *personalised advice* does not have a significant effect on electricity consumption. However, *savings tips* are shown to have a significant positive effect on electricity consumption. This implies that generic savings advice tends to increase electricity consumption rather than reduce it. There is evidence to support hypothesis 2a, and to reject hypothesis 3b. There is inconclusive evidence to neither support nor reject hypothesis 2b and 3a.

Next, it was hypothesised that the use of *social norms* would have a significant reduction effect on residential electricity consumption only in the presence of *injunctive norms*. The results in table 8 tell us that both the use of descriptive *social norms* and *injunctive norms* have a significant negative effect on electricity consumption when other incentives are controlled for. There is evidence to refute hypothesis 4a, *social norms* alone do have the desired effect of reducing electricity demand. There is also evidence to support hypothesis 4b.

Hypothesis 5 refers to the effect of incentives with respect to sample size. We hypothesised that as sample size increases, the effect of an incentive on electricity consumption falls. In fig. 15, we can see that with smaller sample sizes, there is a greater variation in treatment effect, and with larger samples, the treatment effect is smaller. This provides some evidence to support hypothesis 5.

Similarly to previous meta-analyses, we separated the sample set by number of controls used. Higher control studies are those which are deemed to be more representative of the population (use random treatment assignment and an opt-out method of recruitment) and which include greater controls of potential heterogeneity (use a control group, account for weather variation and collect socio-demographic data). Table 4 gives the ATE of high control studies and all other studies. The higher control studies show a statistically significant smaller ATE than the other studies. This provides evidence to support hypothesis 6.

Finally, hypothesis 7 refers to the downward trend in ATE over time.

As can be seen in table 1, previous meta-analyses showed average reduction effects of incentives on electricity consumption of upwards of 6.4%. The present analysis found an overall ATE of 3.37%, or a weighted ATE of 1.85%. This lends support to the hypothesis that the incentives used have a smaller effect on electricity consumption in the Smart Grid Era compared to the eras identified in previous meta-analyses.

7 Conclusion

This paper has provided an analysis of the effects of different incentives in recent electricity conservation studies across the fields of economics, psychology, marketing and building research. This meta-analysis provides the most update assessment of recent experimental literature including newer methods of incentivising consumers to lower their energy consumption.

In conclusion, on average, an incentive designed to reduce household electricity consumption will result in a reduction in consumption of 3.37%. Accounting for the different sized samples used in the individual studies, an incentive can be expected to achieve electricity consumption savings of 1.85%. This result indicates that significant electricity savings can be attained by incentivising households to make behavioural changes to reduce their electricity consumption.

In particular, less costly incentives such as informing households of their individual consumption[9], or of the average consumption in their neighbourhood has a greater negative effect on electricity consumption compared to more costly incentives such as changing pricing for electricity. This has important policy implications given that the latter incentive is often not readily accepted by consumers (Alexander, 2010).

Education campaigns aimed at encouraging households to reduce their consumption via generic energy savings advice, or actions personalised to the household, do not appear to reduce electricity consumption. The former tends to increase consumption.

The results associated with the different control variables show that it is important to undertake field studies which are methodologically rigorous; studies which include control groups, control for demographic information and variations in weather, as well avoiding treatments into which subjects self-select.

The present meta-analyses faces certain limits. To begin with, the meta-analysis is as reliable as the primary studies included in the dataset. Certain

[9] predominately via paper bills or on a website in the current sample of studies

primary studies found treatment effects which were much larger, in both the direction of electricity savings and in consuming more electricity. Such results should not necessarily be excluded from the dataset as the meet the criteria set out in section 4.1, however, they may influence the findings and conclusions of the analysis. Secondly, few experiments test the effect of a single incentive on electricity consumption as they often combine several incentive types. This makes it difficult to separate the effects of individual incentives on electricity consumption due to confounding effects. A third limit concerns the differences in the design of the various studies that are not accounted for in the present study. For example, the composition of the samples in the primary studies is not necessarily identical: participants may have previously participated in similar studies, or the study may focus on a particular type of household.

The findings of this meta-analysis indicate that lower-cost incentives may be sufficient and that there is not necessarily a need to use costly monetary incentives when the objective is to reduce overall electricity consumption. Much focus in recent years has been on injunctive norm based incentives. One conclusion of this analysis is that descriptive social norms may be sufficient on their own.

For future research, this analysis highlights the need to use a control group against which to measure a change in electricity consumption. It is also advisable to consider the methods of treatment assignment and of household selection. These methods should be reflective of the objective being tested. For example, concerning the question of opt-in or opt-out, whether households choose appropriate incentives for them, or whether incentives are applied to large numbers of households. This area merits further research.

References

Mahmoud A Alahmad, Patrick G Wheeler, Avery Schwer, Joshua Eiden, and Adam Brumbaugh. A comparative study of three feedback devices for residential real-time energy monitoring. *Industrial Electronics, IEEE Transactions on*, 59(4):2002–2013, 2012.

Barbara R Alexander. Dynamic pricing? not so fast! a residential consumer perspective. *The Electricity Journal*, 23(6):39–49, 2010.

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

Ian Ayres, Sophie Raseman, and Alice Shih. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, 29(5):992–1022, 2012.

Rene MJ Benders, Rixt Kok, Henri C Moll, Gerwin Wiersma, and Klaas Jan Noorman. New approaches for household energy conservation—in search of personal household energy budgets and energy reduction options. *Energy policy*, 34(18):3612–3622, 2006.

Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. *Introduction to meta-analysis*. Wiley, 2009.

Severin Borenstein, Michael Jaske, and Arthur Rosenfeld. Dynamic pricing, advanced metering, and demand response in electricity markets, 2002.

Jacquelin Burgess and Michael Nye. Re-materialising energy use through transparent monitoring systems. *Energy Policy*, 36(12):4454–4459, 2008.

David Card and Alan B Krueger. Time-series minimum-wage studies: a meta-analysis. *The American Economic Review*, 85(2):238–243, 1995.

James Carroll, Seán Lyons, and Eleanor Denny. Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45:234–243, 2014.

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

Sarah Darby et al. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 486:2006, 2006.

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

Matthias Egger, George Davey Smith, Martin Schneider, and Christoph Minder. Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109):629–634, 1997.

Karen Ehrhardt-Martinez, Kat A Donnelly, Skip Laitner, et al. Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities, 2010.

Torgeir Ericson. Households' self-selection of dynamic electricity tariffs. *Applied Energy*, 88(7):2541–2547, 2011.

Ahmad Faruqui and Sanem Sergici. Arcturus: international evidence on dynamic pricing. *The Electricity Journal*, 26(7):55–65, 2013.

Ahmad Faruqui, Sanem Sergici, and Ahmed Sharif. The impact of informational feedback on energy consumption—a survey of the experimental evidence. *Energy*, 35(4):1598–1608, 2010.

Elisha R Frederiks, Karen Stenner, and Elizabeth V Hobman. Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews*, 41:1385–1394, 2015.

Daphne Geelen, Angèle Reinders, and David Keyson. Empowering the end-user in smart grids: Recommendations for the design of products and services. *Energy Policy*, 61:151–161, 2013.

Maria Gleerup, 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*, pages 113–132, 2010.

Murray Goulden, Ben Bedwell, Stefan Rennick-Egglestone, Tom Rodden, and Alexa Spence. Smart grids, smart users? the role of the user in demand side management. *Energy research & social science*, 2:21–29, 2014.

Alice Grønhøj and John Thøgersen. Feedback on household electricity consumption: learning and social influence processes. *International Journal of Consumer Studies*, 35(2):138–145, 2011.

Vehbi C Gungor, Dilan Sahin, Taskin Kocak, Salih Ergut, Concettina Buccella, Carlo Cecati, and Gerhard P Hancke. Smart grid technologies: Communication technologies and standards. *IEEE transactions on Industrial informatics*, 7(4):529–539, 2011.

Tom Hargreaves, Michael Nye, and Jacquelin Burgess. Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy policy*, 38(10):6111–6119, 2010.

Tim Harries, Ruth Rettie, Matthew Studley, Kevin Burchell, and Simon Chambers. Is social norms marketing effective? a case study in domestic electricity consumption. *European Journal of Marketing*, 47(9):1458–1475, 2013.

Sébastien Houde, Annika Todd, Anant Sudarshan, June A Flora, K Carrie Armel, et al. Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. *Energy Journal*, 34(1):87–102, 2013.

Adnane Kendel and Nathalie Lazaric. The diffusion of smart meters in france: A discussion of the empirical evidence and the implications for smart cities. *Journal of Strategy and Management*, 8(3):231–244, 2015.

Ahsan Raza Khan, Anzar Mahmood, Awais Safdar, Zafar A Khan, and Naveed Ahmed Khan. Load forecasting, dynamic pricing and dsm in smart grid: A review. *Renewable and Sustainable Energy Reviews*, 54:1311–1322, 2016.

Colin McKerracher and Jacopo Torriti. Energy consumption feedback in perspective: integrating australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2):387–405, 2013.

Kenichi Mizobuchi and Kenji Takeuchi. The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in japan. *Energy Policy*, 63:775–787, 2013.

Dean Mountain. Hydro one networks inc.: Tou pricing pilot project results, 2008.

Jon P Nelson and Peter E Kennedy. The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. *Environmental and resource economics*, 42(3):345–377, 2009.

GJ Raw and DI Ross. Energy demand research project: Final analysis, 2011.

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

Frank L Schmidt and John E Hunter. *Methods of meta-analysis: Correcting error and bias in research findings.* Sage publications, 2014.

P Wesley Schultz, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434, 2007.

P Wesley Schultz, Mica Estrada, Joseph Schmitt, Rebecca Sokoloski, and Nilmini Silva-Send. Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms. *Energy*, 90:351–358, 2015.

Maamar Sebri. A meta-analysis of residential water demand studies. *Environment, development and sustainability*, 16(3):499–520, 2014.

TD Stanley, Stephen B Jarrell, and Hristos Doucouliagos. Could it be better to discard 90% of the data? a statistical paradox. *The American Statistician*, 64(1):70–77, 2010.

Tom D Stanley and Hristos Doucouliagos. *Meta-regression analysis in economics and business*, volume 5. Routledge, 2012.

Tom D Stanley and Hristos Doucouliagos. Neither fixed nor random: weighted least squares meta-analysis. *Statistics in medicine*, 34(13):2116–2127, 2015.

Tom D Stanley and Stephen B Jarrell. Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2): 161–170, 1989.

Jonathan AC Sterne, Roger M Harbord, et al. Funnel plots in meta-analysis. *Stata Journal*, 4:127–141, 2004.

Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press, 2008.

Jacopo Torriti. Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in northern italy. *Energy*, 44(1):576–583, 2012.

Tsuyoshi Ueno, Fuminori Sano, Osamu Saeki, and Kiichiro Tsuji. Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data. *Applied Energy*, 83(2):166–183, 2006.

SS Van Dam, CA Bakker, and JDM Van Hal. Home energy monitors: impact over the medium-term. *Building Research & Information*, 38(5):458–469, 2010.

George L Van Houtven, Subhrendu K Pattanayak, Faraz Usmani, and Jui-Chen Yang. What are households willing to pay for improved water access? results from a meta-analysis. *Ecological Economics*, 136:126–135, 2017.

Iana Vassileva, Monica Odlare, Fredrik Wallin, and Erik Dahlquist. The impact of consumers' feedback preferences on domestic electricity consumption. *Applied Energy*, 93:575–582, 2012.

A Included Studies

Author	Year	Publication Information
Alahmad et al.	2012	IEEE Transactions on Industrial Electronics 59(4), 2002-2013
Allcott	2011	Resource and Energy Economics 33(4), 820-842
Allcott	2011	Journal of Public Economics 95(9), 1082-1095
Ayres et al.	2013	Journal of Law, Economics and Organization 29(5), 992-1022
Benders et al.	2006	Energy Policy 34(18), 3612-3622
Carroll et al.	2014	Energy Economics 45, 234-243
Costa and Kahn	2013	Journal of the European Economic Association 11(3), 680-702
Department of Energy & Climate Change	2015	Department of Energy & Climate Change
Dougherty	2013	Opinion Dynamics Corporation
DNV KEMA Energy and Sustainability	2014	DNV KEMA Energy and Sustainability
D'Oca et al.	2014	Energy Research and Social Science 3, 131-142

Author	Year	Publication Information
Faruqui and Sergici	2011	Journal of Regulatory Economics 40(1), 82-109
Gleerup et al.	2010	Energy Journal 113-132
Grønhøj and Thøgersen	2011	International Journal of Consumer Studies 35(2), 138-145
Harries et al.	2013	European Journal of Marketing 47(9), 1458-1475
Houde et al.	2013	Energy Journal 34(1), 87-102
Kendel and Lazaric	2015	Journal of Strategy and Management 8(3), 231-244
Kua and Wong	2012	Energy Policy 47, 49-56
Martin and Rivers	2015	(working paper)
Mizobuchi and Takeuchi	2013	Energy Policy 63, 775-787
Mountain	2006	Hydro One Network Inc.
Mountain	2008	Hydro One Networks Inc.
Mountain	2012	Research Institute for Quantitative Studies in Economics and Population

Author	Year	Publication Information
Nilsson at al.	2014	Applied Energy 122, 17-23
Parker et al.	2008	Florida Solar Energy Center
Provencher et al.	2015	Navigant
Raw and Ross	2011	Energy Demand Research Project: Final Analysis
Schleich et al.	2013	Energy Policy 61, 1097-1106
Schultz et al.	2015	Energy 90, 351-358
Schumatz and Dimetrosky	2014	NMR Group Inc and Tetra Tech
Shen et al.	2016	Energy Policy 98, 19-32
Sullivan et al.	2013	Freeman, Sullivan & Co.
Sullivan et al.	2016	Nexant
Torriti	2012	Energy 44(1), 576-583
Ueno et al.	2006	Applied Energy 83(2), 166-183

Author	Year	Publication Information
Van Dam et al.	2010	Building Research and Information 38(5), 458-469
Van Elburg	2014	Dutch Energy Savings Monitor for the Smart Meter
Vassileva et al.	2012	Applied Energy 93, 575-582
Xu et al.	2015	Energy Procedia 75, 2694-2699

Table 9: Studies included in analysis

B Reasons for Study Exclusion

Reason for exclusion	Number of papers excluded
Different treatment effect measure	28
(e.g.: peak demand reduction, appliance level data, median % change)	
Non-residential sample	17
Not a field experiment or pilot study	11
(e.g.: a simulated study or laboratory experiment)	
Included under a different title	8
Gas and electricity consumption combined	4
Experimental issues leading to missing data	3
Secondary data	1
Total	72

Table 10: Reasons for excluding studies from the analysis

C Cluster-robust OLS Model

Table 11 displays the results of the cluster-robust OLS meta-regression model with estimations separated by incentive type as baseline. The first specification includes all variables, specifications 2-4 focus on a particular incentive strategy (monetary, personal feedback and social feedback) and the final specification includes only the study design variables. Each estimation includes a variable accounting for the duration of the study and the type of literature it is from. Finally, standard errors for each estimation are clustered by study to account for non-independence between studies. Coefficients on the different incentives are interpreted as the increase or decrease in percentage points of electricity consumption due to a particular incentive relative to the consumption of the control group, when present in the study, which is the case for 90% of the observations, or the baseline level of consumption. A negative coefficient signifies a reduction in electricity consumption by the number of percentage points displayed.

	(1) Full	(2) Monetary	(3) Personal feedback	(4) Social feedback	(5) Study design
Monetary incentive	1.651	2.075			
	(1.055)	(1.354)			
Monetary information	1.980	-0.333			
	(1.523)	(1.009)			
Individual feedback	-2.116		-1.754		
	(1.273)		(1.260)		
Real-time feedback	-2.565*		-1.253		
	(1.310)		(1.048)		
Personalised advice	-2.252		-1.690		
	(2.069)		(1.993)		
Savings tips	1.455		1.195		
	(1.292)		(1.338)		
Social norms	-3.462*			-1.161	
	(1.976)			(1.875)	
Injunctive norms	-1.942			-3.113	
	(2.557)			(2.825)	
Control group	5.586*	3.420	4.983	4.070	3.499
	(2.854)	(3.131)	(3.116)	(3.026)	(3.201)
Weather controls	-0.778	-0.003	0.294	0.380	0.405
	(1.339)	(1.053)	(1.095)	(1.342)	(1.182)
Demographic controls	0.094	-1.640	-1.665	-1.576	-2.419
	(1.603)	(1.623)	(1.706)	(1.802)	(1.901)
Random assignment	-1.116	-0.189	-0.343	0.154	0.135
	(1.299)	(0.990)	(1.083)	(1.302)	(1.049)
Opt-in recruitment	-4.224	-1.441	-1.652	-3.063	-0.847
	(2.799)	(1.495)	(1.470)	(2.739)	(1.038)
Duration	0.215***	0.144**	0.164***	0.135***	0.117***
	(0.056)	(0.058)	(0.038)	(0.045)	(0.043)
Peer reviewed	3.603**	1.982*	2.302**	2.301*	1.601
	(1.391)	(1.112)	(1.106)	(1.146)	(0.971)
Constant	-7.455*	-7.597**	-7.410	-6.166	-6.732*
	(4.016)	(3.395)	(4.423)	(4.082)	(3.765)
Observations	105	105	105	105	105
Adjusted R^2	0.194	0.132	0.157	0.114	0.109

Standard errors in parentheses
Standard errors are clustered by primary study.
A negative coefficient reads as a reduction in energy consumption.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: OLS estimation of treatment effects