

Khanna, Tarun M.; Bruns, Stephan; Miersch, Klaas; Minx, Jan C.

Working Paper

Effects of electricity pricing schemes on household energy consumption. A meta-analysis of academic and non-academic literature

Suggested Citation: Khanna, Tarun M.; Bruns, Stephan; Miersch, Klaas; Minx, Jan C. (2023) : Effects of electricity pricing schemes on household energy consumption. A meta-analysis of academic and non-academic literature, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/270889>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Appendix

1.1. Literature search and selection data extraction

Our data collection strategy involved (1) a review of relevant existing reviews and the studies identified by them; (2) string-based searches of bibliographic databases; and (3) searches for grey literature on Google.

We screened the studies referenced by existing reviews (Buckley, 2020; Delmas et al., 2013; Faruqui et al., 2013, 2017; Faruqui & Sergici, 2010; Khanna et al., 2021; Srivastava et al., 2018). We searched for articles that dealt with dynamic pricing of energy or electricity for households (see Appendix

Table 1) for inclusion-exclusion criteria). We only tagged as relevant studies that dealt with energy consumption by households or dormitories and contained a quantitative estimate for the energy saved through a relevant intervention. Studies where no obvious comparator group was available (untreated control group or pre-intervention data). Studies that only measured energy consumption of certain appliances or direct load control of appliances by utilities, with no possibility for the household to intervene were not considered.

In addition to the literature identified above, in accordance with guidance for rigorous evidence syntheses, we searched a broad set of bibliographic databases (Web of Science Core Collections Citation Indexes, Scopus, JSTOR), the working paper repository RePEc, and the web-based academic search engine Google Scholar, based on a comprehensive search string. The Web of Science Core Collection Citation Indexes included in our search were: Science Citation Index Expanded (SCI-EXPANDED) --1900-present, Social Sciences Citation Index (SSCI) --1900-present, Arts & Humanities Citation Index (A&HCI) --1975-present, Conference Proceedings Citation Index- Science (CPCI-S) --1990-present, Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-present, Emerging Sources Citation Index (ESCI) --2015-present. We developed a search string that followed the PICOS (population, intervention, comparator, outcome, and study design) logic recommended by the Campbell Collaboration (Appendix Table 2). The search string was developed iteratively by checking the results of the search against a set of studies of known relevance. Since we did not make any exclusions based on the date, methodology or the field of publication, the searches returned a large number of studies (50,698) after removing duplicates. To enable screening of relevant papers, we applied a machine learning algorithm using support vector machines (M. W. Callaghan & Müller-Hansen, 2020) to rank the studies in the order of relevance of their abstracts. Article screening was done first at the title and abstract level, assisted by machine learning, and then manually on full text level. A training database of relevant studies was derived from previous reviews and known studies. This database was used to train the machine learning model that analyzed the abstracts of the studies to rank them in the order of predicted relevance. We conducted an

iterative process where at each iteration, we 1) trained the algorithm with the already screened documents, 2) fitted this enhanced model on the unseen documents and 3) assigned the next set of documents for review by selecting the documents predicted to be most relevant. In total, we screened 3,571 studies at abstract level, which allowed for a recall of 75% at $p = 0.064$. Out of these, 261 documents were rated relevant. We applied a revised stopping criterion (M. W. Callaghan & Müller-Hansen, 2020) for abstract screening using a biased urn adjustment with a conservative assumed bias = 10, which allowed for 85% work savings.

In addition to academic research, many electric power utilities, especially in the US, have implemented pilot programs to reduce consumers' electricity consumption during peak hours. These programs are regularly evaluated, and results published in the form of annual reports and not published in academic journals. Therefore, in addition to searching academic literature, we manually searched for utility reports which are not commonly included in academic databases. We conducted an internet search starting with the list of utilities provided by Faruqui et al. (2017) and also identified five main databases that contain relevant reports, the California Measurement Advisory Council (CALMAC), The Department of Energy's Office of Electricity (OE) Resource Library, Public Utilities Commission California, Association of Energy Services Professionals (AESP) Resource Library, and Illinois Commerce Commission (ICC) Database. We then searched these databases for relevant policy evaluation reports.

In total we found 453 studies, which is more than twice the number of studies included in the existing reviews. Unlike other reviews, we further restricted our analysis to studies in which complete quantitative details, including an estimate for the monetary incentive size. A team of four reviewers then manually screened the abstracts of the top 3571 studies. Full text screening was performed on a selection of 453 studies deemed relevant from this initial screening. The final sample included 85 studies after critical appraisal. The ROSES flowchart for screening and coding is shown in Appendix Figure 1) and the complete list of studies included in the analysis is available in Appendix

Table 7 and **Error! Reference source not found.**).

Four reviewers extracted the relevant data from these studies using the rules laid out in a codebook (available in the supplementary material on Github). To ensure consistency, a sample of 50 studies was screened at the abstract level between any two reviewers. The reviewers next did a full text screening and coded the relevant papers from this sample, followed by discussion of the coded fields to see what disagreements occurred and made suitable adjustments to the codebook.

Table 1: PICOS framework summary of inclusion and exclusion criteria

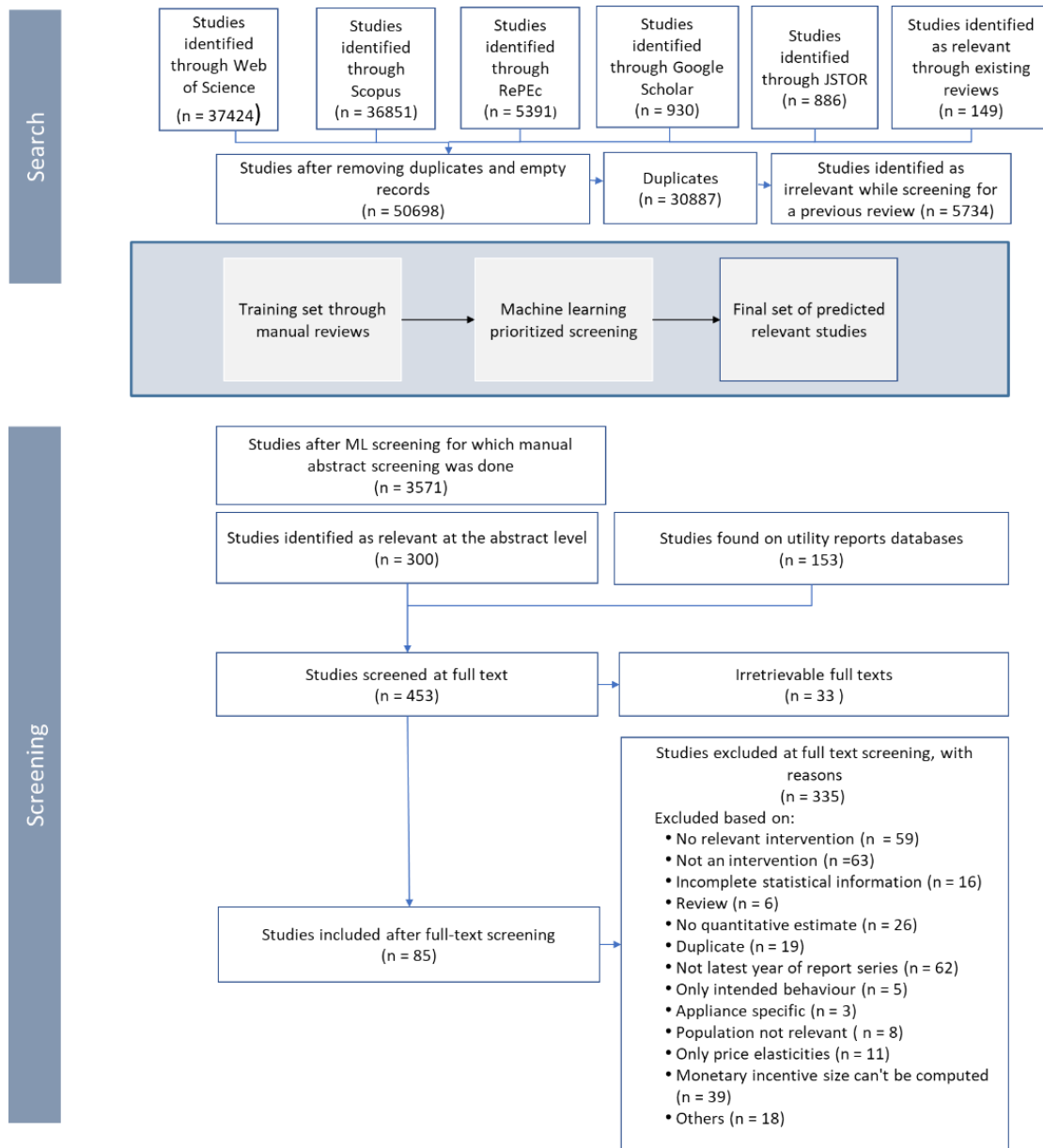
Dimension	Inclusion	Exclusion
Population	Households globally (including residential dormitories)	Industrial, Commercial buildings
Intervention	Monetary incentives for individuals/households to reduce residential electricity and heat consumption	Behavioural interventions without monetary incentive, direct load control without behavioural component
Comparator	Households without the intervention, before the intervention began	Without comparator
Outcome	Observed electricity and heat consumption	Purely hypothetical studies, outcomes unrelated to energy consumption as response to pricing
Study Type	Empirical quantitative studies	Simulation, modelling or predictive studies

Table 2: Full Boolean Search String used on Web of Science, Scopus, JSTOR and RePEc

	Search Keywords
Population	((household* OR residential OR building OR dormitor* OR individual OR consumer* OR participant* OR customer* OR domestic OR homeowner*))
Intervention	(pric* OR "time-of-use" OR "time-of-day" OR "real time" OR "peak" OR "dynamic pricing" OR "smart meter*" OR "smart grid*" OR (behavioral AND (economic* OR intervention* OR guideline*))) OR rebate OR reward OR incentives OR tariff OR "time-varying pricing")
Comparator	-
Outcome	((energy OR electric* OR gas) NEAR/14 (consumption OR conservation OR efficiency OR use OR demand OR usage)) OR "price responsiveness")
Study type	-

Note: The exact syntax of the presented search string does only work on Web of Science. For the other databases the syntax was adopted according to the platform requirements. As Google scholar does not allow for long search queries the following query was used: (price OR "time of use" OR peak OR tariff OR rebate OR reward) AND (household* OR residential) AND ("electricity consumption" OR "energy consumption" OR "energy conservation" OR shifting).

Figure 1: ROSES flowchart for search and screening of the relevant literature



1.2. Robustness check for using effective sample size as a measure for precision

The studies in our sample did not always consistently report standard errors along with the estimates. In order to estimate the precision of each estimate, we, therefore, rely on the effective sample size for each observation. We estimate the standard error for each observation using the following formula:

$$\text{standard error (effective)} = 1 / \frac{(\text{treatment sample size} * \text{control sample size})}{(\text{treatment sample size} + \text{control sample size})}$$

The robustness of using this estimator was checked on a sub-set of the studies for which both the standard error reported by studies and the effective standard error as calculated above were available and that reported the effect sizes with the same dependent variable. Figure 2 shows the comparison of precision calculated using reported and effective standard error. The meta-analysis for the peak consumption using both estimates of precision (Figure 3) shows that the results from using the effective standard error are comparable to those from using the standard error reported by the studies. There is higher variation in the estimates from the fixed effects estimation but in the case of all the estimates the 95% confidence intervals overlap and the results are qualitatively similar. Therefore, we conclude that the use of effective sample size and the use of the same results as the use of standard errors.

Figure 2 Comparison of precision calculated using reported and effective standard error

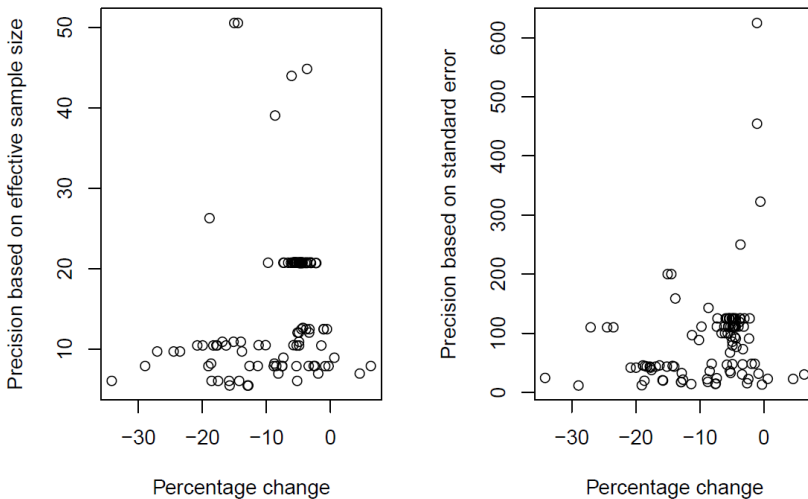


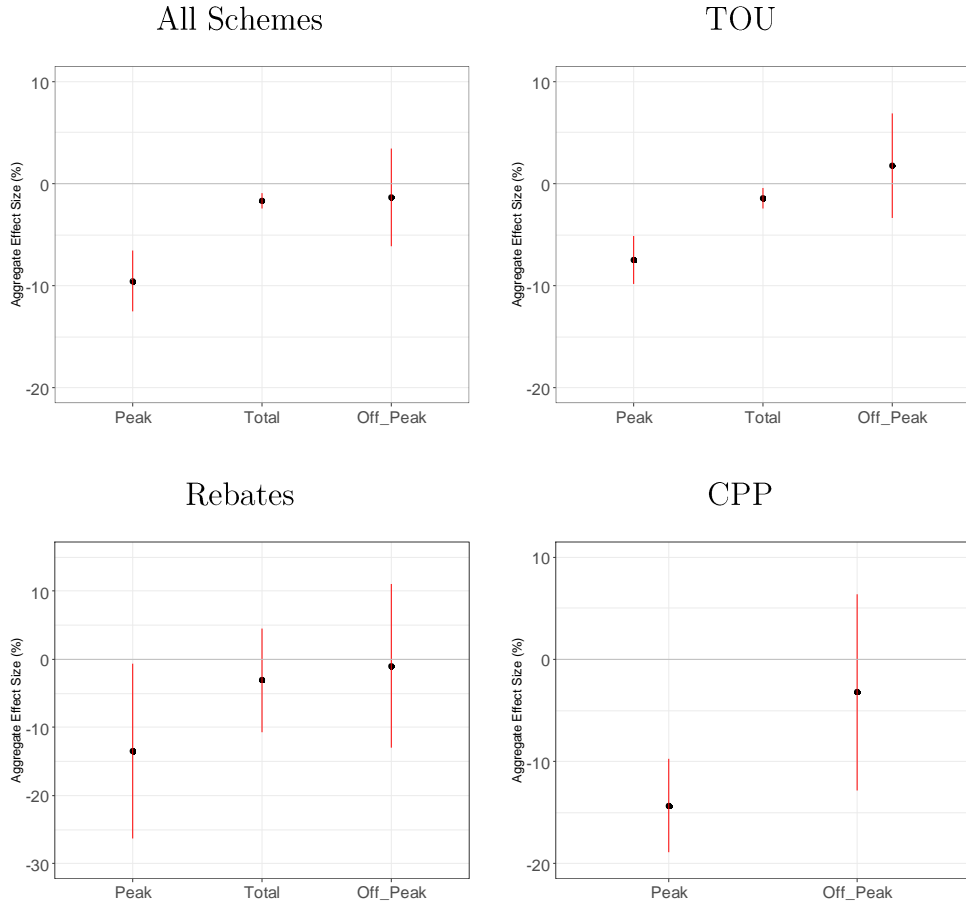
Figure 3 Results for average treatment effect for peak consumption with various models calculated using reported and effective standard error

	RE	FE	RE-PET	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-10.617 *** (1.703)	-4.566 . (2.375)					-11.255 . (2.596)	-9.629 (4.573)
intercept			-6.107 ** (2.191)	-2.100 (2.886)	-8.751 *** (1.923)	-4.245 (2.421)		
se (effective)			-196.140 ** (19.981)	-413.242 . (226.767)				
v (effective)					-2080.747 ** (275.946)	-4797.014 (4225.304)		

	RE	FE	RE-PET	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-10.617 *** (1.703)	-7.754 ** (2.213)					-11.255 . (2.596)	-10.709 (3.569)
intercept			-2.257 (4.009)	-7.921 . (4.012)	-6.270 ** (2.518)	-7.382 * (2.801)		
se			-90.792 . (32.965)	3.439 (48.610)				
v					-414.939 . (134.536)	-115.104 (223.383)		

1.3. Robustness checks: results from fixed effects model

Figure 4: Aggregate effect sizes across interventions and the type of demand measured.



Notes: Results from the FE model with standard errors clustered at the study level

Table 3: Estimates of publication bias and 95% confidence intervals for the average treatment effect by type of demand (results from the fixed-effects model)

	Total Consumption				Peak Consumption			
	FE	FE-PET	FE-PEESE	FE-Top10	FE	FE-PET	FE-PEESE	FE-Top10
overall	-1.7*			-1.7*	-9.7*			-10.5*
	[-2.4; -1.0]			[-2.4; -1.0]	[-12.2; -7.2]			[-13.3; -7.7]
intercept		-1.9*	-1.7*			-8.3*	-9.1*	
		[-2.9; -0.9]	[-2.3; -1.0]			[-12.8; -3.8]	[-12.4; -5.7]	
ser: Peer-Reviewed Paper		-1.6				-11.3		
		[-23.0; 19.8]				[-81.4; 58.9]		

	Total Consumption				Peak Consumption			
	FE	FE-PET	FE-PEESE	FE-Top10	FE	FE-PET	FE-PEESE	FE-Top10
ser: Unpublished Academic Paper		-56.8* [-69.9; -43.6]				-12.2 [-89.1; 64.8]		
ser: Utility Report		7.3 [-13.1; 27.7]				-42.0 [-106.1; 22.1]		
ver: Peer-Reviewed Paper			-38.0 [-130.9; 55.0]				-173.1 [-616.0; 269.8]	
ver: Unpublished Academic Paper			-484.5* [-575.4; -393.5]				30.3 [-304.6; 365.1]	
ver: Utility Report			36.5 [-89.3; 162.2]				-226.4 [-629.1; 176.3]	
DF Resid.	128	125	125	113	331	328	328	32
nobs	129	129	129	114	332	332	332	33

Notes: PET = Precision effect test and PEESE = precision-effect estimate with SE. Models estimated using the fixed effects model using the *metafor* package in R with robust standard errors clustered at the study level. The square brackets report 95% confidence intervals obtained from bootstrap clustering. The star (*) denotes that the null hypothesis value (zero) is outside the confidence interval.

1.4. Moderator variables for effect heterogeneity

The meta-regression models that investigate the causes for heterogeneity in effect sizes were estimated using the random effects and multilevel models and introducing moderator variables in the estimation equation. Mathematically, the interpretation of the parameter on a moderator variable in meta-regression is the same as for a parameter estimate from a traditional regression; that is, it represents the average change in the effect size associated with one-unit change in the moderator. Moderator variables could represent factors that genuinely affect the magnitude of the relationship between the focal predictor and the outcome of interest or could represent design elements of original studies that may affect effect size from coded studies (Ringquist 2013). In this study we include both type of moderator variables. Design elements of original

studies are captured as dummy variables, for example, for the following variables: controls (if a study controls for demographic variables like age, income, composition of the family etc., it is assigned value 1); residence type (if a study controls for the characteristics of the house like size etc., it is assigned value 1). We also include as moderator variables study design (difference-in-difference, control-treatment, or pre-post) and statistical method (panel regression, OLS regression, or difference of means tests) employed in the studies. Other moderator variables capture the factors that are likely to affect the relationship between energy use and the treatment, for example, duration of experiment or region in which the experiment was performed. The complete list of the moderator variables, along with the summary statistics, included in the heterogeneity analysis is shown in Table 4.

Table 4: Summary statistics of the coded study characteristics

No	Variable	Stats / Values	Freqs (% of Valid)
1	PercentageChange [numeric]	Mean (sd) : -6.4 min ≤ med ≤ max: -38.8 ≤ -4.8 ≤ 43 IQR (CV) : 9.5 (-	495 distinct values
2	Peak [numeric]	Min : 0 Mean : 0.6 Max : 1	0 : 239 (43.00%) 1 : 317 (57.00%)
3	Mid_Peak [numeric]	Min : 0 Mean : 0 Max : 1	0 : 531 (95.50%) 1 : 25 (4.50%)
4	Total [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 456 (82.00%) 1 : 100 (18.00%)
5	Off_Peak [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 442 (79.50%) 1 : 114 (20.50%)
6	CPP [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 419 (75.40%) 1 : 137 (24.60%)
7	RTP [numeric]	Min : 0 Mean : 0 Max : 1	0 : 550 (98.90%) 1 : 6 (1.10%)
8	TOU [numeric]	Min : 0 Mean : 0.6 Max : 1	0 : 217 (39.00%) 1 : 339 (61.00%)
9	Rebates [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 482 (86.70%) 1 : 74 (13.30%)
10	ser [numeric]	Mean (sd) : 0.1 (0) min ≤ med ≤ max: 0 ≤ 0.1 ≤ 0.3 IQR (CV) : 0.1	145 distinct values
11	UtilityReport [numeric]	Min : 0 Mean : 0.6	0 : 229 (41.20%) 1 : 327 (58.80%)

No	Variable	Stats / Values	Freqs (% of Valid)
		Max : 1	
12	Unpublished [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 528 (95.00%) 1 : 28 (5.00%)
13	PeerReviewed [numeric]	Min : 0 Mean : 0.4 Max : 1	0 : 355 (63.80%) 1 : 201 (36.20%)
14	MonetaryIncentiveSize [numeric]	Mean (sd) : 4.6 min ≤ med ≤ max: 1 ≤ 3.5 ≤ 25 IOR (CV) : 3 (0.7)	133 distinct values
15	anyEnablingTechnology [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 373 (67.10%) 1 : 183 (32.90%)
16	anyEnergySavingsFeedback [numeric]	Min : 0 Mean : 0.4 Max : 1	0 : 343 (61.70%) 1 : 213 (38.30%)
17	OLS [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 444 (79.90%) 1 : 112 (20.10%)
18	MeansDifferences [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 434 (78.10%) 1 : 122 (21.90%)
19	Other_stats [numeric]	Min : 0 Mean : 0 Max : 1	0 : 553 (99.50%) 1 : 3 (0.50%)
20	PanelEffects [numeric]	Min : 0 Mean : 0.5 Max : 1	0 : 276 (49.60%) 1 : 280 (50.40%)
21	ControlTreatment [numeric]	Min : 0 Mean : 0.4 Max : 1	0 : 334 (60.10%) 1 : 222 (39.90%)
22	PrePost [numeric]	Min : 0 Mean : 0 Max : 1	0 : 551 (99.10%) 1 : 5 (0.90%)
23	DID [numeric]	Min : 0 Mean : 0.6 Max : 1	0 : 227 (40.80%) 1 : 329 (59.20%)
24	Asia [numeric]	Min : 0 Mean : 0 Max : 1	0 : 534 (96.00%) 1 : 22 (4.00%)
25	Europe [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 476 (85.60%) 1 : 80 (14.40%)
26	Canada [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 450 (80.90%) 1 : 106 (19.10%)
27	UnitedStates [numeric]	Min : 0 Mean : 0.6 Max : 1	0 : 232 (41.70%) 1 : 324 (58.30%)
28	Others [numeric]	1 distinct value	0 : 556 (100.00%)
29	InterventionDuration [numeric]	Mean (sd) : 42.1 min ≤ med ≤ max: 0 ≤ 41 ≤ 192 IOR (CV) : 48	39 distinct values
30	PeakDuration [numeric]	Mean (sd) : 4.9 min ≤ med ≤ max: 0 ≤ 5 ≤ 16 IOR (CV) : 3 (0.7)	12 distinct values
		Min : 0	0 : 382 (68.70%)

No	Variable	Stats / Values	Freqs (% of Valid)
31	OptedIn [numeric]	Mean : 0.3 Max : 1	1 : 174 (31.30%)
32	StudyYear [numeric]	Mean (sd) : 2010.7 min ≤ med ≤ max: 1979 ≤ 2014 ≤ IOR (CV) : 10 (0)	20 distinct values
33	HouseholdType [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 469 (84.40%) 1 : 87 (15.60%)
34	ResidenceType [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 510 (91.70%) 1 : 46 (8.30%)
35	Season [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 375 (67.40%) 1 : 181 (32.60%)
36	Weather [numeric]	Min : 0 Mean : 0.4 Max : 1	0 : 324 (58.30%) 1 : 232 (41.70%)
37	Heating [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 478 (86.00%) 1 : 78 (14.00%)
38	Cooling [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 384 (69.10%) 1 : 172 (30.90%)
39	HeatingCooling [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 445 (80.00%) 1 : 111 (20.00%)
40	OutBias [numeric]	Min : 0 Mean : 0.6 Max : 1	0 : 197 (35.40%) 1 : 359 (64.60%)
41	ReportBias [numeric]	Min : 0 Mean : 0.1 Max : 1	0 : 496 (89.20%) 1 : 60 (10.80%)
42	LogMIPeak [numeric]	Mean (sd) : 0.8 min ≤ med ≤ max: 0 ≤ 0.9 ≤ 3.2 IOR (CV) : 1.4 (1)	109 distinct values
43	LogMITotal [numeric]	Mean (sd) : 0.2 min ≤ med ≤ max: 0 ≤ 0 ≤ 2.5 IOR (CV) : 0 (2.3)	50 distinct values

1.5. Robustness checks for the Bayesian model averaging

To address these issues, we employ the Bayesian model averaging technique (BMA)—an approach used in previous meta-analysis in economics by Bajzik et al., 2020; and Havranek et al., 2012, 2017). In the baseline specification, we employ the unit information g-prior that the regression coefficient is zero the same weight as one observation of the data (Eicher et al., 2011). This agnostic prior reflects our lack of knowledge regarding the probability of individual parameter values. To test the robustness of our estimates we use the dilution prior that adjusts

model probabilities by multiplying them by the determinant of the correlation matrix of the variables included in the model (Havranek et al., 2012, 2017). Furthermore, as another robustness check, we follow Fernandez et al. (2012) who use the so-called BRIC g-prior. The results from these alternative priors are presented below.

Table 5: Results for the Bayesian Meta Analyses (BMA) with different priors for total consumption

Variable	UIP-Uniform			UIP-Dilution			BRIC-g		
	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD
CPP	0.08	- 0.06	0.29	0.03	- 0.03	0.19	0.03	- 0.03	0.18
TOU	0.17	0.19	0.50	0.05	0.05	0.26	0.04	0.04	0.24
ser	0.06	- 0.53	2.85	0.02	- 0.12	1.37	0.01	- 0.10	1.24
UtilityReport	0.04	- 0.00	0.13	0.01	- 0.00	0.07	0.01	- 0.00	0.06
EnablingTechnology	1.00	- 4.78	0.68	1.00	- 4.74	0.65	1.00	- 4.74	0.65
EnergySavingsFeed back	0.08	0.09	0.37	0.02	0.02	0.17	0.02	0.02	0.16
OLS	0.37	- 0.66	0.99	0.13	- 0.22	0.63	0.11	- 0.19	0.58
MeansDifferences	0.07	- 0.07	0.36	0.02	- 0.01	0.13	0.01	- 0.01	0.11
ControlTreatment	0.04	- 0.01	0.14	0.01	- 0.00	0.07	0.01	- 0.00	0.07
Asia	0.63	- 2.34	2.09	0.35	- 1.33	1.98	0.32	- 1.22	1.93
Europe	0.05	0.03	0.24	0.02	0.01	0.15	0.02	0.01	0.14
Canada	0.04	0.02	0.19	0.02	0.01	0.13	0.01	0.01	0.12

Variable	UIP-Uniform			UIP-Dilution			BRIC-g		
	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD
InterventionDuration	0.99	0.04	0.01	0.98	0.04	0.01	0.98	0.04	0.01
PeakDuration	0.03	- 0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.01
OptedIn	0.87	- 1.69	0.87	0.65	- 1.26	1.05	0.62	- 1.21	1.06
StudyYear	0.05	0.00	0.01	0.02	0.00	0.00	0.01	0.00	0.00
HouseholdType	0.04	- 0.02	0.20	0.01	- 0.01	0.10	0.01	- 0.00	0.09
ResidenceType	0.13	0.22	0.68	0.07	0.15	0.62	0.07	0.15	0.61
Season	0.06	0.03	0.25	0.03	0.02	0.18	0.02	0.02	0.18
Weather	0.51	0.83	0.92	0.54	0.97	0.99	0.54	0.98	1.00
Heating	0.05	0.03	0.25	0.02	0.02	0.20	0.02	0.02	0.20
Cooling	0.19	- 0.23	0.54	0.08	- 0.10	0.38	0.07	- 0.09	0.36
HeatingCooling	0.04	0.00	0.15	0.01	0.00	0.08	0.01	0.00	0.07
OutBias	0.09	0.09	0.34	0.02	0.02	0.17	0.02	0.02	0.15
ReportBias	0.04	0.03	0.26	0.02	0.01	0.16	0.01	0.01	0.15
LogMIPeak	1.00	- 6.37	0.33	1.00	- 6.40	0.33	1.00	- 6.40	0.33
LogMITotal	0.05	- 0.02	0.15	0.01	- 0.01	0.08	0.01	- 0.00	0.07
(Intercept)	1.00	- 3.89	NA	1.00	- 2.06	NA	1.00	- 1.96	NA

Table 6: Results for the Bayesian Meta Analyses (BMA) with different priors for peak consumption

Variable	UIP-Uniform			UIP-Dilution			BRIC-g		
	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD
CPP	0.18	0.38	1.00	0.04	0.06	0.44	0.02	0.02	0.28
TOU	0.68	2.28	1.97	0.27	0.80	1.49	0.18	0.51	1.24
ser	0.33	-8.49	13.81	0.26	-6.72	12.39	0.21	-5.33	11.2
UtilityReport	0.04	0.00	0.20	0.02	0.00	0.11	0.01	0.00	0.08
anyEnablingTechnology	1.00	-4.10	1.06	0.99	-3.71	1.08	0.99	-3.62	1.07
anyEnergySavingsFeedback	0.04	0.01	0.24	0.02	0.01	0.18	0.01	0.01	0.15
OLS	0.21	-0.46	1.05	0.05	-0.10	0.51	0.03	-0.04	0.34
MeansDifferences	0.06	-0.01	0.41	0.02	-0.01	0.22	0.01	-0.01	0.17
ControlTreatment	0.18	-0.33	0.83	0.10	-0.18	0.62	0.07	-0.12	0.51
Asia	0.90	-5.45	2.54	0.90	-6.04	2.59	0.91	-6.28	2.53
Europe	0.06	-0.08	0.49	0.02	-0.02	0.24	0.01	-0.01	0.18
Canada	0.04	0.02	0.27	0.02	0.01	0.17	0.01	0.01	0.14
InterventionDuration	0.11	0.00	0.01	0.04	0.00	0.00	0.03	0.00	0.00
PeakDuration	0.12	0.03	0.09	0.06	0.01	0.06	0.04	0.01	0.05

Variable	UIP-Uniform			UIP-Dilution			BRIC-g		
	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD
OptedIn	0.05	-0.03	0.26	0.03	-0.03	0.21	0.02	-0.02	0.18
StudyYear	0.40	0.05	0.08	0.11	0.01	0.04	0.06	0.01	0.03
HouseholdType	0.32	-1.02	1.72	0.05	-0.11	0.61	0.02	-0.03	0.33
ResidenceType	0.36	1.45	2.25	0.15	0.57	1.52	0.10	0.39	1.28
Season	0.15	0.25	0.72	0.11	0.22	0.68	0.10	0.19	0.65
Weather	0.19	0.32	0.76	0.11	0.19	0.60	0.09	0.15	0.54
Heating	0.04	0.01	0.24	0.02	-0.00	0.15	0.01	-0.00	0.12
Cooling	0.70	-2.10	1.64	0.89	-3.16	1.40	0.92	-3.36	1.29
HeatingCooling	0.11	0.23	0.81	0.02	0.03	0.27	0.01	0.01	0.18
OutBias	0.18	-0.33	0.84	0.06	-0.09	0.42	0.04	-0.06	0.32
ReportBias	0.73	3.76	2.81	0.29	1.38	2.43	0.18	0.89	2.08
LogMI	1.00	-4.64	0.80	1.00	-5.07	0.74	1.00	-5.17	0.70
(Intercept)	1.00	-109.70	NA	1.00	-25.26	NA	1.00	-13.25	NA

Table 7: Complete list of academic studies included in the analysis

	Authors	Year	Title
1	Aigner, D.; Lillard, L.	1984	Measuring Peak Load Pricing Response from Experimental Data: An Exploratory Analysis
2	Allcott, H	2011	Rethinking real-time electricity pricing
3	Andersen, L.M.; Hansen, L.G.; Jensen, C.L.; Wolak, F.A.	2019	Can Incentives to Increase Electricity Use Reduce the Cost of Integrating Renewable Resources
4	Bartusch, C	2011	Wester, Introducing a demand based electricity distribution tariff in the residential sector: demand response and customer perception,
5	Battalio, R.; Kagel, J.; Winkler, R.; Winett, R.	1979	Residential Electricity Demand: An Experimental Study
6	Belenguer, E; Garcia, N; Sabater-Grande, G	2019	Assessment of energy efficiency improvement methods in the residential sector through the development of economic experiments
7	Burkhardt, J.; Gillingham, K.; Kopalle, P.K.	2019	Experimental Evidence on the Effect of Information and Pricing on Residential Electricity Consumption
8	Capitan, T; Alpizar, F; Madrigal-Ballesteros, R; Pattanayak, SK; Pattanayak, SK	2021	Time-varying pricing may increase total electricity consumption: Evidence from Costa Rica
9	Carroll, J; Lyons, S; Denny, E	2014	Reducing household electricity demand through smart metering: The role of improved information about energy saving
10	Dolan, P	2013	Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation (Economics Working Paper No. 2589269).
11	Eguiguren-Cosmelli, JM	2018	Responsiveness of low-income households to hybrid price/non-price policies in the presence of energy shortages: evidence from Colombia
12	Faruqui, A	2011	Dynamic pricing of electricity in the mid-Atlantic region: econometric results from the Baltimore gas and electric company experiment,
13	Fowle, M.; Wolfram, C.; Spurlock, A.C.; Todd, A.; Baylis, P.	2017	Default Effects and Follow-On Behavior: Evidence from an Electricity Pricing Program
14	Fowle, M; Fowle, M; Wolfram, C; Wolfram, C; Baylis, P; Spurlock, CA; Todd-Blick, A; Cappers, P	2021	Default Effects And Follow-On Behaviour: Evidence From An Electricity Pricing Program
15	Garnache, C; Imenes, A; Oystein, Hernaes	2022	Which Households Respond to Electricity Peak Pricing Amid High Levels of Electrification?
16	Geelen, D; Keyson, D; Boess, S; ...	2012	Exploring the use of a game to stimulate energy saving in households

	Authors	Year	Title
17	Harold, J; Lyons, S; Cullinan, J	2018	Heterogeneity and persistence in the effect of demand side management stimuli on residential gas consumption
18	Herter, K; McAuliffe, P; Rosenfeld, A	2007	An exploratory analysis of California residential customer response to critical peak pricing of electricity
19	Herter, K; Wayland, S	2010	Residential response to critical-peak pricing of electricity: California evidence
20	Herter, K; Wood, V; Blozis, S	2013	The effects of combining dynamic pricing, AC load control, and real-time energy feedback: SMUD'S 2011 Residential Summer Solutions Study
21	Ho, T.T.; Shinkuma, S.; Shimada, K.	2018	The effects of dynamic pricing of electric power on consumer behavior: A propensity score analysis for empirical study on Nushima Island, Japan
22	Hofmann, M; Hofmann, M; Lindberg, KB; Lindberg, KB	2021	Do households react to variable power prices? - Results from a Norwegian pricing experiment
23	Ito, K., Ida, T., & Tanaka, M.	2018	Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand
24	Jessoe, K; Rapson, D	2014	Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use
25	Jessoe, K; Rapson, D; Smith, JB	2014	Towards understanding the role of price in residential electricity choices: Evidence from a natural experiment
26	Matsukawa, I.	2018	Information acquisition and residential electricity consumption: Evidence from a field experiment
27	MIDDEN, CJH; METER, JE; WEENIG, MH; ZIEVERINK, HJA	1983	USING FEEDBACK, REINFORCEMENT AND INFORMATION TO REDUCE ENERGY-CONSUMPTION IN HOUSEHOLDS - A FIELD-EXPERIMENT
28	Mizobuchi, K; Takeuchi, K	2013	The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in Japan
29	Mizutani, F	2015	The Effect of Demand Response on Electricity Consumption in Japan
30	Murakami, Kayo; Shimada, H.; Ushifusa, Y.; Ida, T.	2022	Heterogeneous Treatment Effects of Nudge and Rebate: Causal Machine Learning in a Field Experiment on Electricity Conservation
31	Nguyen, TTK; Shimada, K; Ochi, Y; Matsumoto, T; Matsugi, H; Awata, T	2016	An Experimental Study of the Impact of Dynamic Electricity Pricing on Consumer Behavior: An Analysis for a Remote Island in Japan
32	Pon, S	2017	The Effect of Information on TOU Electricity Use: An Irish Residential Study
33	Sudarshan, A	2017	Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives
34	Suter, JF; Shammin, MR	2013	Returns to residential energy efficiency and conservation measures: A field experiment

	Authors	Year	Title
35	Taylor, T.; Schwarz, P.	1990	The Long-Run Effects of a Time-of-Use Demand Charge
36	Torriti, J	2012	Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy
37	Wenjie Wang and Takanori Ida	2014	Default Effect Versus Active Decision: Evidence from a Field Experiment in Los Alamos
38	Wolak, F	2006	Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment
39	Wolak, FA	2011	Do residential customers respond to hourly prices? Evidence from a dynamic pricing experiment
40	Woo, C.K. Horowitz, I & Sulyima, I	2013	Relative kW response to residential time-varying pricing in British Columbia

Table 8: Complete list of utility reports included in the analysis (in descending order by publication year)

	Authors/Consultant	Utility/Grid Operator	Year	Title
1	-/RLW Analytics	AmrenUE	2004	AmerenUE Residential TOU Pilot Study Load Research Analysis First Look Results
2	-/Charles River Associates	California statewide pricing pilot	2005	Impact evaluation of the California statewide pricing pilot.
3	-/RLW Analytics	Idaho Power	2006	2006 Analysis at the Residential Time-at-Day and Energy Watch Pilot Programs Final Report. Submitted to the Idaho Public Utilities Commission, Case No. IPC-E-06-05
4	Puckett, H; Hennessy T/RLW Analytics	AmrenUE	2006	Ameren UE Residential TOU Pilot Study Load Research Analysis - 2005 Program Results
5	-/RLW Analytics	AmrenUE	2006	Time-of-Day and Energy Watch Pilot Programs Final Report
6	Strapp, C/IBM Global Business Services and eMeter Strategic Consulting	Ontario Energy Board	2007	Ontario Energy Board Smart Price Pilot Final Report. Prepared by IBM Global Business Services and eMeter Strategic Consulting for the Ontario Energy Board
7	Hammerstrom, D/-	Pacific Northwest National Laboratory	2007	Pacific Northwest Grid Wise Test bed Demonstration Projects, Part1. Olympic Peninsula Project. Pacific Northwest National Laboratory Report Number PNNL- 17167, prepared for US Department of Energy under contract DE-AC05-76L01830.
8	-/ Summit Blue Consulting	PSE&G	2007	Final report for the Mypower Pricing Segments Evaluation.
9	-/Summit Blue	CNT Energy	2007	Evaluation of the 2006 Energy-Smart Pricing Plan Final Report. Prepared for CNT Energy
10	-/Hydro One Networks	Hydro One Networks	2008	Time-of-Use Pricing Pilot Project Results. Report Number EB-2007-0086 delivered to the Ontario Energy Board
11	-/Navigant Consulting Inc.	Newmarket Hydro Ltd	2008	EVALUATION OF TIME-OF-USE PRICING PILOT
12	Faruqui, Ahmad; Sergici, Sanem/The Brattle Group	Baltimore Gas & Electric Company	2009	BGE's SMART ENERGY PRICING PILOT SUMMER 2008 IMPACT EVALUATION

13	George, Stephen S et al/Freeman, Sullivan & Co	Pacific Gas & Electric Company	2010	2009 Load Impact Evaluation for Pacific Gas and Electric Company's Residential SmartRate™—Peak Day Pricing and TOU Tariffs and SmartAC Program Volume 1: Ex Post Load Impacts Final Report
14	Gary Martin/-	Electric Ireland	2011	Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report
15	-/Navigant	Ameren Illinois Utilities	2011	Power Smart Pricing 2010 Annual Report
16	-/Global Energy Partners	Oklahoma Gas & Electric Company (OGE)	2011	OG&E SMART STUDY TOGETHER IMPACT RESULTS Interim Report Summer 2010
17	Hansen et al/Electric Power Research Institute	First Energy	2013	FirstEnergy's Consumer Behavior Study: Preliminary Evaluation for the Summer 2012
18	Schare, Stuart et al./Navigant Consulting Inc.	NSTAR	2013	NSTAR Smart Grid Pilot Technical Performance Report #1 AMR Based Dynamic Pricing DE-OE0000292
19	George, S.; Jimenez, L; Potter, J/-	Sacramento Municipal Utility	2014	SmartPricing Options Final Evaluation
20	-/DTE Energy	DTE Electric	2014	Final Evaluation Report
21	Karen Herter, Ph.D. Yevgeniya Okuneva/Herter Energy Research Solutions,	Sacramento Municipal Utility District	2014	SMUD's 2013 PowerStat Study Final Report
22	-/Nexant	San Diego Gas & Electric Company	2014	2013 Load Impact Evaluation of San Diego Gas & Electric Company's Opt-in Peak Time Rebate Program
23	-/Navigant	Commonwealth Edison Company	2015	ComEd Residential Real-Time Pricing (RRTP) Program Annual Report 2014
24	Blumsack, Seth; Hines, Paul /-	Green Mountain Power	2015	Load Impact Analysis of Green Mountain Power Critical Peak Events, 2012 and 2013
25	Faruqui, A., Lessem, N., Sergici, S., Mountain, D., Denton, F., Spencer, B., & King, C./The Brattle Group	Independent Electric System Operator	2016	Analysis of Ontario's Full Scale Roll-out of TOU Rates—Final Study
26	Keene, A. et al/Seventhwave Minnesota Power	Minnesota Power	2016	Minnesota Power's Advanced Metering Infrastructure Project AMI Behavioral Research Pilot—Phase Two: Results from the Time-of-Day Rate with Critical Peak Pricing Pilot Program

27	-/Navigant	CustomerFirst-ESQ Pricing for 3 LDC's (Greater Sudbury Hydro)	2020	CustomerFirst Regulated Price Plan Pilot Program: Final Report - Final Impact and Process Evaluation Report
28	Sergici, S., Faruqui, A. et al./Brattle Group	Maryland Joint Utilities (Baltimore Gas & Electric, Pepco Maryland, Delmarva Power& Light Maryland)	2020	PC44 Time of Use Pilots: Year One Evaluation
29	Sergici, S., Faruqui, A. et al./Brattle Group	Baltimore Gas & Electric	2020	PC44 Time of Use Pilots: Year One Evaluation
30	Sergici, S., Faruqui, A. et al./Brattle Group	Pepco Maryland	2020	PC44 Time of Use Pilots: Year One Evaluation
31	Sergici, S., Faruqui, A. et al./Brattle Group	Delmarva Power& Light Maryland	2020	PC44 Time of Use Pilots: Year One Evaluation
32	-/Publicis Sapient	Oshawa Power	2020	RPP PILOT PROGRAM – FINAL RESULTS REPORT
33	Neetika Sathe, Daniel Carr, Shirley Liu/-	Alectra Utilities	2021	Alectra Utilities Advantage Regulated Price Plan Pilot – Final Report
34	-/Nexant	San Diego Gas & Electric Company	2021	2020 Load Impact Evaluation of San Diego Gas & Electric's Default Time-of-Use (TOU) Rates
35	-/Christensen Associates Energy Consulting	San Diego Gas & Electric Company	2021	2020 Load Impact Evaluation of San Diego Gas & Electric's Voluntary Residential Critical Peak Pricing (CPP) and Time-of-Use (TOU) Rates
36	Bell et al. /Nexant	Southern California Edison	2021	2020 Load Impact Evaluation of Southern California Edison's Default Time-of-Use Pilot

Table 9: Publication bias for sub-sets of data: a) peak consumption, academic literature b) peak consumption, utility reports c) total consumption, academic literature d) total consumption, utility reports

a								
	RE	FE	RE-pet	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-10.4*	-9.7*					-10.3*	-9.7*
	[-12.1; -8.7]	[-12.2; -7.2]					[-12.1; -8.4]	[-12.2; -7.2]
intercept			-7.4*	-8.4*	-9.5*	-9.1*		
			[-11.0; -3.9]	[-13.2; -3.7]	[-11.8; -7.3]	[-12.4; -5.8]		
ser			-32.3	-25.1				

			[-67.1; 2.4]	[-79.7; 29.5]				
ver					-82.6	-167.9		
					[-249.1; 84.0]	[-452.2; 116.3]		

* Null hypothesis value outside the confidence interval.

b

	RE	FE	RE-pet	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-10.0*	-11.0*					-10.2*	-11.0*
	[-12.5; -7.6]	[-13.2; -8.7]					[-13.0; -7.4]	[-13.3; -8.7]
intercept			-9.0*	-10.6*	-9.7*	-10.9*		
			[-12.5; -5.6]	[-16.2; -5.0]	[-12.4; -7.0]	[-13.8; -7.9]		
ser			-10.3	-7.1				
			[-31.4; 10.9]	[-81.3; 67.1]				
ver					-27.8	-33.5		
					[-64.6; 9.0]	[-335.3; 268.2]		

* Null hypothesis value outside the confidence interval.

c

	RE	FE	RE-pet	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-4.1*	-2.1					-2.0	-2.0
	[-7.4; -0.8]	[-4.4; 0.2]					[-5.4; 1.4]	[-4.3; 0.4]
intercept			0.8	-2.0	-0.8	-1.8		
			[-3.3; 4.9]	[-5.6; 1.6]	[-4.2; 2.7]	[-4.5; 1.0]		
ser			-30.6*	-0.7				
			[-56.6; -4.6]	[-32.2; 30.8]				
ver					-97.6*	-33.3		
					[-179.6; -15.6]	[-137.2; 70.7]		

* Null hypothesis value outside the confidence interval.

d

	RE	FE	RE-pet	FE-PET	RE-PEESE	FE-PEESE	RE-Top10	FE-Top10
overall	-2.6*	-1.7*					-1.5*	-1.7*
	[-4.2; -1.0]	[-2.4; -1.0]					[-3.0; -0.1]	[-2.4; -1.0]
intercept			-0.3	-1.5*	-1.2	-1.5*		
			[-2.3; 1.6]	[-2.4; -0.5]	[-2.6; 0.2]	[-2.2; -0.9]		
ser			-19.3	-4.5				

			[-39.5; 0.9]	[-23.0; 14.0]				
ver					-72.4*	-43.5		
					[-137.4; - 7.5]	[-125.1; 38.2]		
* Null hypothesis value outside the confidence interval.								

References

- Callaghan, Max, Finn Müller-Hansen, Jérôme Hilaire, and Yuan Ting Lee. 2020. “NACSOS: NLP Assisted Classification, Synthesis and Online Screening (Version v0.1.0).”
- Collaboration for Environmental Evidence. 2013. *Guidelines for Systematic Review and Evidence Synthesis in Environmental Management Version 4.2*.
- Haddaway, Neal Robert, and Biljana Macura. 2018. “The Role of Reporting Standards in Producing Robust Literature Reviews.” *Nature Climate Change*.
- Ringquist, Evan J. 2013. *Meta-Analysis for Public Management and Policy*. San Francisco: John Wiley & Sons.
- Viechtbauer, Wolfgang. 2010. “Conducting Meta-Analysis in R with Metafor Package.” *Journal of statistical software* 36(3): 1–48.

