日本評価学会　社会実験分科会　2022年度研究報告会

公共政策におけるエビデンス活用の課題と解決の方向性：シロン※

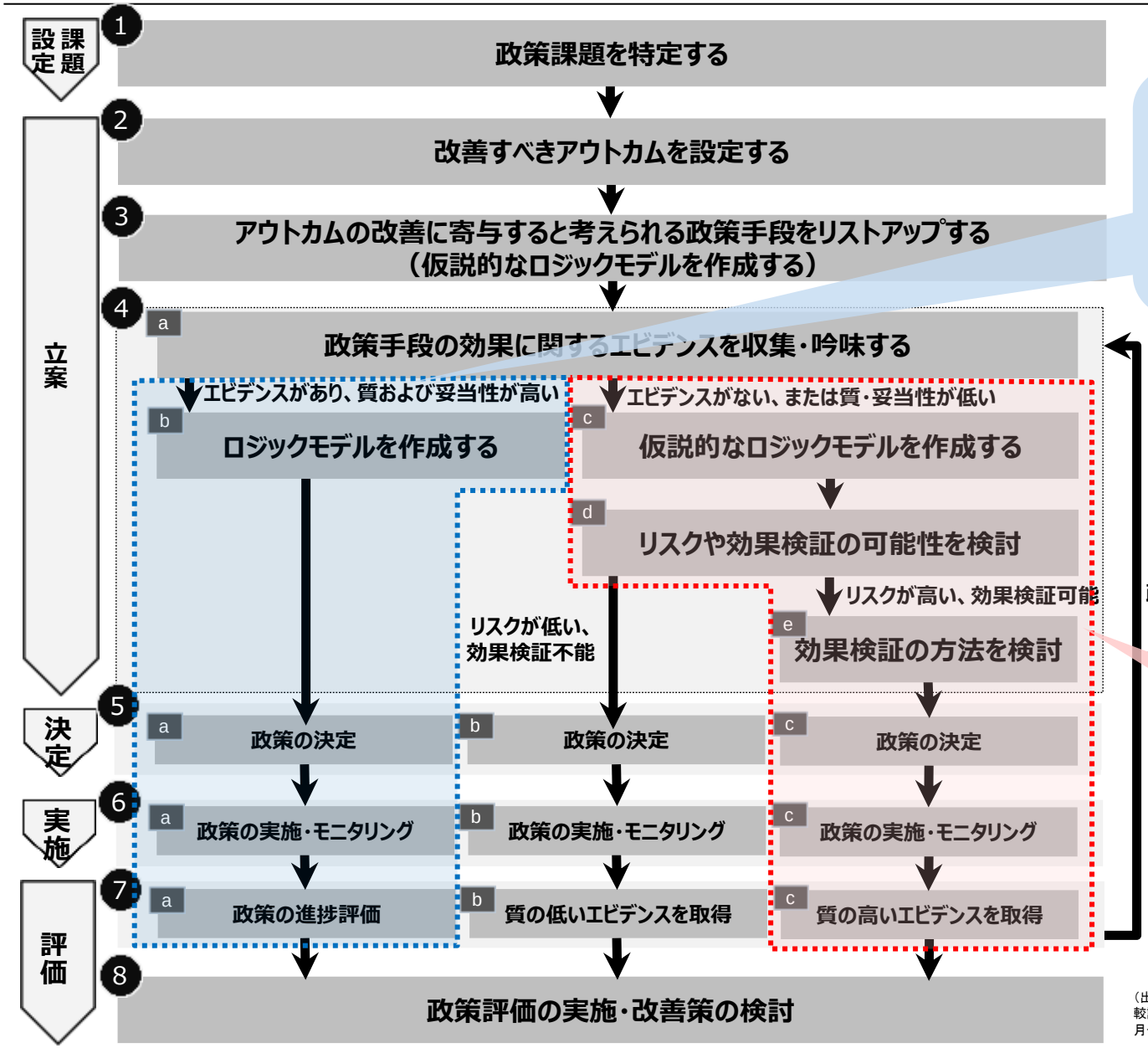
（※）私論・試論

2022年6月25日

三菱UFJリサーチ&コンサルティング株式会社
経済政策部 兼 行動科学チーム（MERIT）　小林　庸平

公共政策におけるエビデンスの活用の重要性と Evidence-Based Practiceにおける 「つくる」「つたえる」「つかう」

Evidence-Basedな公共政策立案と「つくる」「つかう」



良いエビデンスがあるのであれば、EBPMは簡単（つかう）

既存のエビデンスをできるだけ活用したい

エビデンスが乏しく、効果検証が必要な場合は手間暇がかかる（つくる）

（出所）青柳恵太郎・小林庸平（2019）「EBPMの思考法 やってみようランダム化比較試験！vol.4 ランダム化比較試験の実施工程(1)」『経済セミナー』2019年10・11月号を一部加筆修正。

エビデンスを参照する枠組み：評価設問

■エビデンスを参照する場合、はじめに評価設問を立てることが多い。
■評価設問に答える既存のエビデンスが存在するのかを探索することになる。

	評価設問の要素	具体例
P Patient/ Population	**誰に対して**	妊婦に対して
I Intervention	**どんな介入を行い**	コロナワクチンを接種させ
C Comparison	**誰と比較して**	未接種の妊婦と比較して
O Outcome	**何に対する効果を**	コロナ罹患率・重症化率／胎児への影響

EBMの推進に関する３つのステップ：「つくる」「つたえる」「つかう」

■EBMでは、それを社会全体で推進する場合、エビデンスを「つくる」「つたえる」「つかう」という３つのステップが重要だとされている。

つくる	ランダム化比較試験などを行い効果を検証する
つたえる	システマティック・レビューなどを行い、 エビデンスの質の評価と統合を行う
つかう	医師や行政官等が意思決定する

（出所）中山健夫（2010）「エビデンス：つくる・伝える・使う」『体力科学』59, pp.259-268、正木朋也・津谷喜一郎（2006）「エビデンスに基づく医療（EBM）の系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来」『日本評価研究』第6巻第1号, pp.3-19を参考に作成

公共政策における
「つくる」「つたえる」「つかう」の
実現可能性

公共政策における評価設問の実際

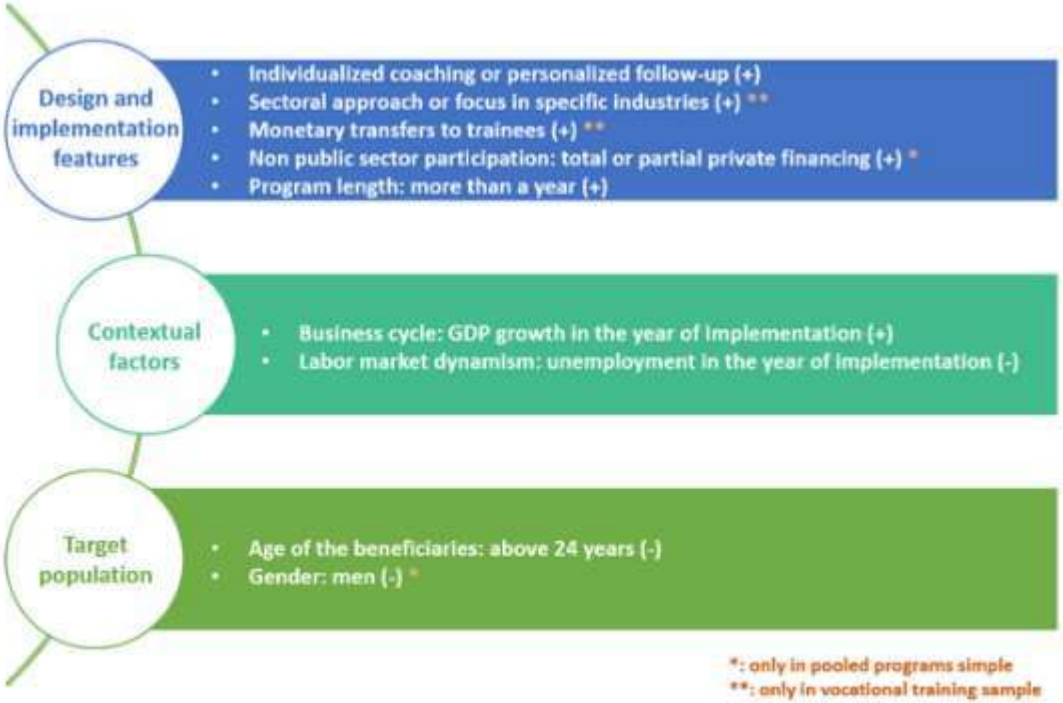
		公共政策における質的変化	具体例
P Patient /Population	誰に対して	課題をかかえている人たちが必ずしも自明ではない 企業や地域なども含まれる	孤立、長期失業 中小企業
I Intervention	どんな介入を行い		
C Comparison	誰と比較して		
O Outcome	何に対する効果を	アウトカムが多元的になりやすい 測定の難しいアウトカムも多い	
要素の拡大 T Time	いつ	好況期なのか・不況期なのか、製造業主体の時代なのかデジタル時代なのか、 （Pの細分化ともいえる）	
要素の拡大 L Location	どこで	都市なのか地方なのか、国内なのか国外なのか （Pの細分化ともいえる）	
要素の拡大 D Delivery	どうデリバリーするか	誰が介入を実施するのか（行政官なのか、民間の委託事業者なのか）、 介入実施者の質、スケールアップ時の質の劣化（Iの細分化ともいえる）	

積極的労働市場政策の効果

Yeyati et al. (2019) What Works for Active Labor Market Policies?

積極的労働市場政策に関する102のRCT、652の推定結果を比較した結果、以下のような結論が得られれた。

- 1/3の推定結果は、統計的に有意かつプラスの効果が確認されている。
  【効果量の異質性が大きい】
- 経済成長率が高く、失業率が低い局面において、プログラムはプラスの効果を得やすい。【Time】
- 職業訓練や賃金補助といった人的資本蓄積を目的したプログラムは、大きな効果を示しやすい。
- プログラム期間、金銭的なインセンティブ、個別フォローアップ、産業ターゲティングなどは、介入の有効性を決定する重要な要素である。【Delivery】



（出所）Yeyati et al,（2019）“Understanding what works for active labour market policies” VOX EU

問題提起

■EBMで想定される「つくる」「つたえる」「つかう」というプロセスは、エビデンスを十分に貯めたうえで、活用可能性を高められれば、意思決定の質を向上できるという考え方が背景にあると考えられる。

■それはEBPMにおいても論理的には正しいと考えられるが、評価設問の要素の質的変化や多元化によって、エビデンスを貯めたとしても意思決定の質の向上にはなかなかつながらないのではないか。

■加えて、公共政策では、政策現場における効果検証実施の調整コストや、ランダム化比較試験の適用可能性の低さから、エビデンスの創出力に限界がある。

■言い換えると、公共政策におけるエビデンスは、外的妥当性が低く文脈依存性が高いく、かつエビデンス創出力が低くなるため、システマティック・レビューのような方法で「つたえる」努力をしたとしても、意思決定に寄与しにくいのではないか。

公共政策の効果検証においてランダム化比較試験が適用可能な範囲

効果検証における要素				分析手法の例
SUTVA	分離	ｻﾝﾌﾟﾙｻｲｽﾞ	介入への関与	
○	○	○	○	・ RCT（ランダム化比較試験） ・ 前向きRDD（回帰不連続デザイン）
○	○	○	×	・ 自然実験、DID（差の差分析） ・ 部分識別
○	○	×	○or×	・ 合成コントロール法
○	×	○	×	・ 構造推定による税制シミュレーション分析 ・ 育児休業延長のシミュレーション分析
×	○	○or×	○or×	・ 一般均衡モデル
×	○	×	×	・ 構造推定によるシミュレーション分析
×	×	×	×	・ 動学的一般均衡モデルによるマクロ分析 ・ 時系列モデル
				・ インタビュー ・ アンケート調査

RCTが適用可能なのはこれらの要素がそろったときのみ

（出所）小林庸平（2020）「日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題－Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのかー」『日本評価研究』第20巻2号pp.33-48

公共政策におけるエビデンス活用に関するシロン

公共政策におけるエビデンス活用に関するシロン

① 問いの設定や事前分析にコストをかける

- EBPMでは、妥当な評価設問すら自明ではないケースが少なくない。
- 評価設問が妥当でないと、効果検証の「無駄打ち」が発生するが、公共政策におけるエビデンス創出力の低さを踏まえると、無駄打ちはコストが大きい。
- であれば、問い（評価設問）の設定や事前分析にコストをかけることが正当化されるのではないか。（因果のエビデンスだけに拘泥しない方が良いのではないか)

② 「つくる」「つたえる」「つかう」にこだわりすぎない

- EBPMでは、評価設問の要素の質的変化や多元化が生じるため、得られるエビデンスは、外的妥当性が低く文脈依存性が高い。
- 「つくる」「つたえる」「つかう」というプロセスを回すことは簡単ではないのではないか。

③ オーダーメイドのエビデンスをつくり、エビデンスの質に関するリスクと向き合う

- EBPMでは、エビデンスを踏まえることで政策的意思決定のリスクを減じ、意思決定を改善できれば良い。
- エビデンスの質をどこまで求めるかは、当該政策の重要性次第となる。
- エビデンスの質は低かったとしても、そのリスクに対して向き合いながらも、オーダーメイドなエビデンスのほうが意思決定の改善に寄与するのであれば、それを活用していけば良いのではないか。
- エビデンスが使われるかどうかは、その質よりも介入の導入しやすさに依存する。

ナッジに関するエビデンス採用のボトルネック

DellaVigna et al. (2022) Bottlenecks for Evidence Adoption

アメリカの67都市における73のナッジのRCTの結果から、エビデンスの採用状況をフォローアップし、ボトルネックを検証した。

- 他の分野と比較して、ナッジ介入のエビデンスは採用されやすく、27%が活用されていた。
- エビデンスが採用されるかどうかは、エビデンスの強さとはあまり関係ない。
- RCT実施時の職員がその業務を引き続き担当していたかどうかは、エビデンスの採用に大きな影響を与える。
- エビデンスの採用にもっとも影響を与えるのは、既存のコミュニケーション（ナッジの活用場面）のなかで行われたかどうか。エビデンスの採用において新しいコミュニーションが必要だと、時間や予算が余計にかかってしまう。

（出所）DellaVigna, Kim, and Linos（2022）“Bottleneck for Evidence Adoption” NBER Working Paper 30144

公共政策におけるエビデンス活用に関するシロン（つづき）

④ 低コストで効果検証できるインフラを構築する

- テック企業は日々大量のRCTを実施。
- 効果検証の障壁は、（特に対照群の）データ取得と、介入コスト（時間的、金銭的）。
- データインフラを整備して、デジタルツールを使ったスピーディな介入が可能になれば、手低コストで効果検証できるようになる。

⑤ 機械学習との融合を進め、異質な介入効果の推定を行う

- 技術的な進展によって、HTE（Heterogeneous Treatment Effect）やCATE（Conditional Average Treatment Effect）が推定できるようになってきた。
- これがより進展すれば、評価設問の要素の質的変化や多元化があったとしても、まとめてそれらに答えられる可能性がある。
- 「つくる」プロセスをスピードアップできる可能性。

⑥ 介入効果の異質性の理論的な解明

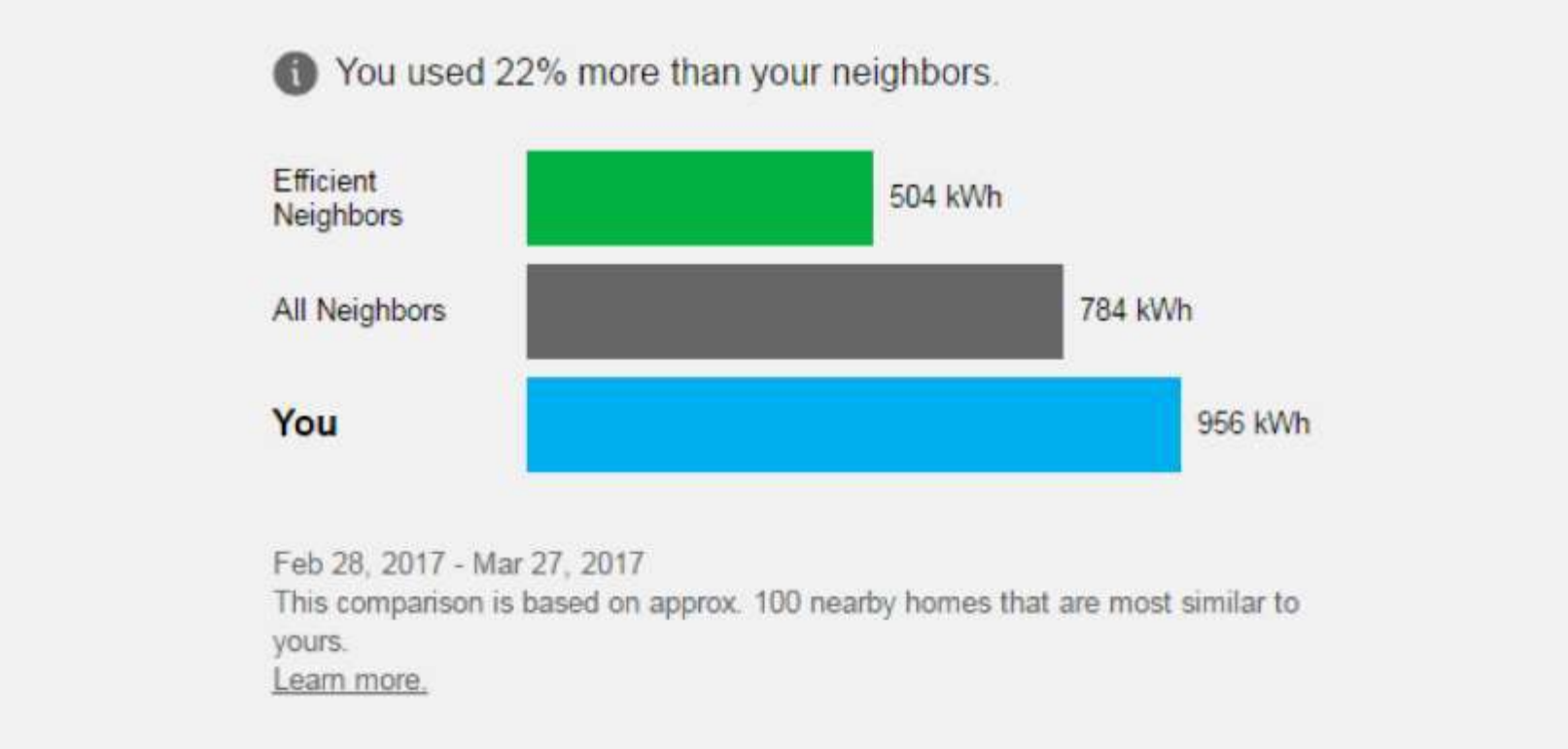
- HTEやCATEが蓄積されると、効果に異質性が生じる背景に理論的な説明を加えられる可能性がある。
- 理論的な説明が加えられれば、エビデンスの一般化可能性や説明可能性、移転可能性を高められる。
- その結果、既存のエビデンスを行政官自らの意思決定に適用し得るかどうかを判断できる。

家庭用エネルギーレポート

■節電を促すために、社会的比較を加味した以下のような家庭用エネルギーレポート（Home Energy Report：HER）を送付。

■HERによって、平均で１か月あたり8.85kWH（1%）の節電が達成される

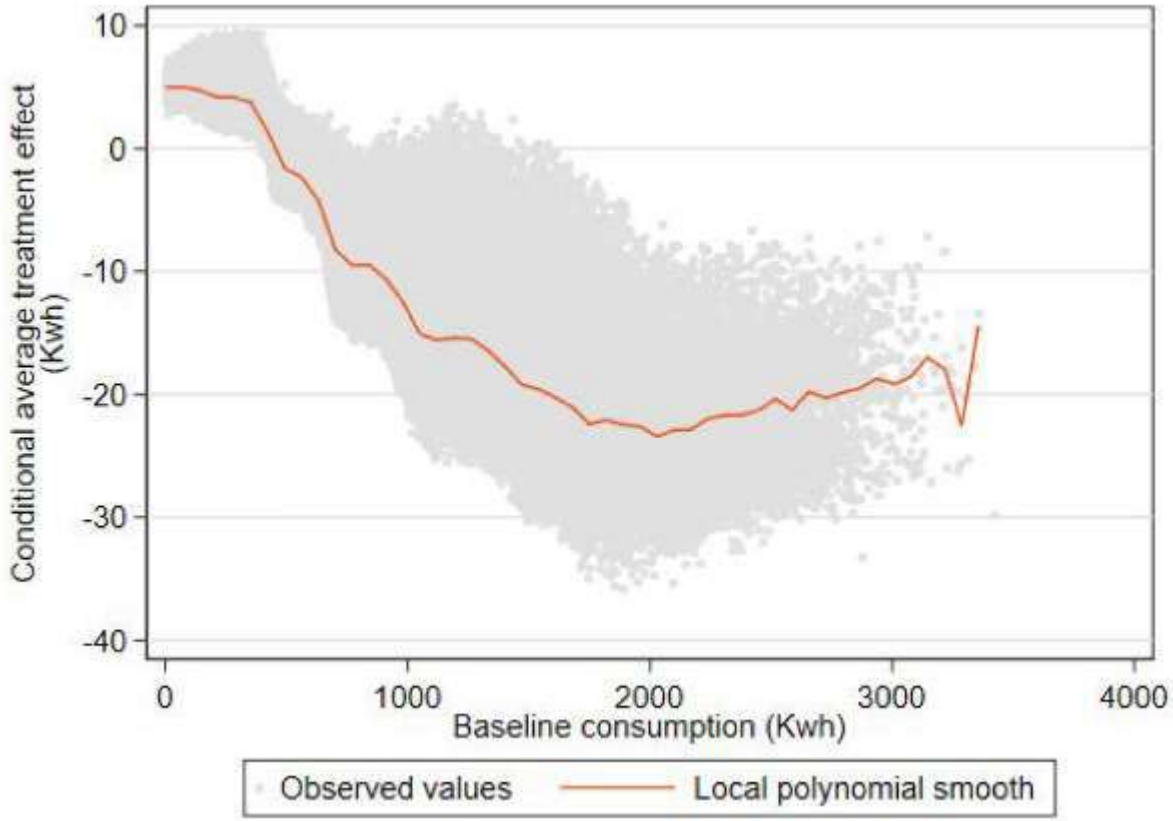
家庭用エネルギーレポートのイメージ



（出所）Knittel, C. and Stolper, S.（2019）"Using Machine Learning to Target Treatment: The Case of Household Energy Use" NBER Working Paper Series 26531

世帯によって異なる家庭用エネルギーレポート

世帯によって異なる家庭用エネルギーレポートの効果



（出所）Knittel, C. and Stolper, S.（2019）"Using Machine Learning to Target Treatment: The Case of Household Energy Use" NBER Working Paper Series 26531

■家庭用エネルギーレポートによる節電効果は、もともとの世帯の電力消費量によって大きな差がある。

■電力消費量の小さな世帯には効果が小さいが、大きな世帯では20kWhの節電効果。