What's the right method to find how much energy smart meters save?

March 24, 2021
15h00 – 16h00

Leonardo ENERGY Webinar Channel
j.mp/leonardotube

16th Webinar of the UsersTCP Academy
www.userstcp.org

UsersTCP Leonardo ENERGY European Copper Institute Copper Alliance

"How to smartly measure energy savings from smart meters."

Speakers:
Andrew Schein, BIT
Kevin Gornall, BEIS

Smart meters are a critical part of the energy transition, but how much energy does their installation save? Measuring savings from smart meters is not easy. How do we model smart metered households' counterfactual consumption? How much energy would these households have consumed had their supplier not installed a smart meter?

What's the right method to find how much energy smart meters save?

UsersTCP 24 March 2021

Andrew Schein

Senior Advisor, Energy and Sustainability,
Behavioural Insights Team (BIT)

Kevin Gornall

Principal Research Officer at Department for
Business, Energy and Industrial Strategy (BEIS)

Introductions

THE BEHAVIOURAL INSIGHTS TEAM

Why did BEIS commission this guidance?

The GB smart meter roll-out - 53 million smart electricity and gas meters

UK Government policy but delivered by energy suppliers

Household energy consumption reductions a key target, delivered via:

- Direct feedback and energy management technologies (e.g. In-Home Energy Displays)
- Indirect feedback (e.g. accurate bills)
- Advice and guidance
- Motivational campaigns (e.g. Smart Energy GB)

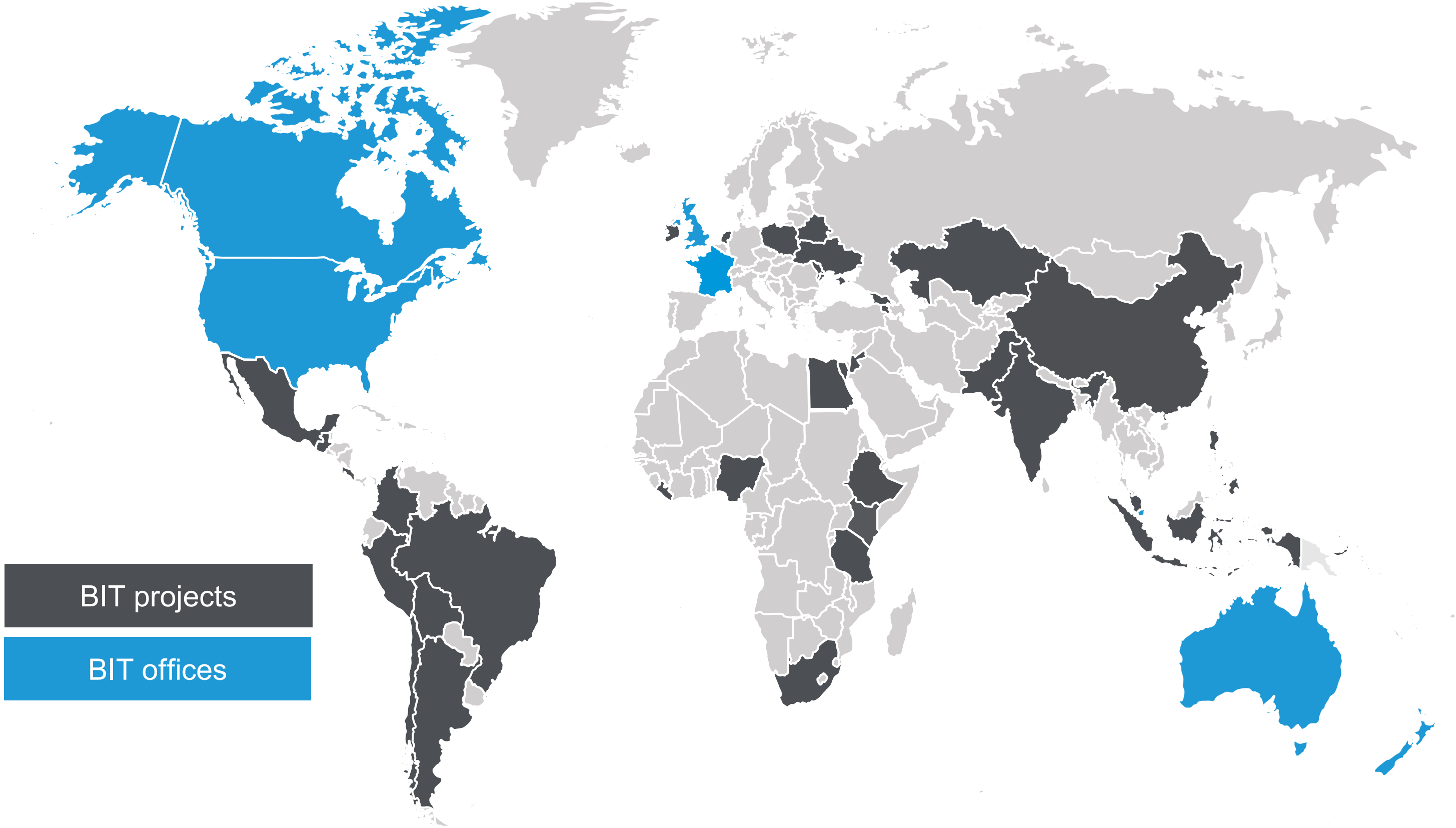
Hypothesis: sustained 3% reduction in electricity consumption, 2.2% reduction in gas consumption

‘Tactical’ quasi-experimental analyses crucial in our monitoring and evaluation strategy for this research question

Guidance designed for energy suppliers - many are conducting this analysis for their own interest

BEIS does not have direct access to the necessary data, so this analysis is greatly appreciated

A BIT of background

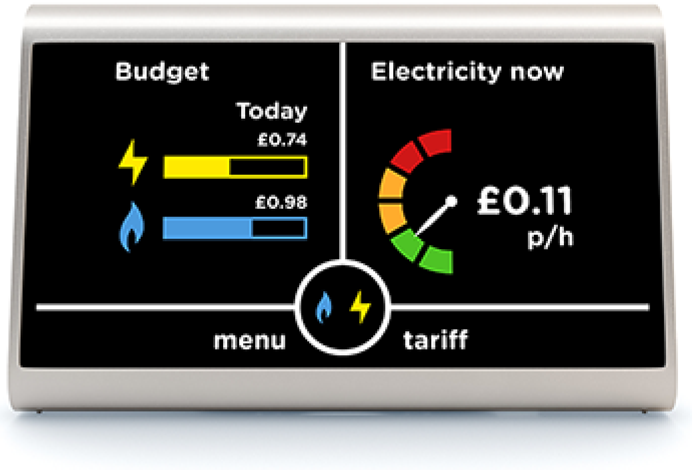


Energy consumption analysis

BIT developed guidance for energy suppliers on behalf of the Department for Business, Energy and Industrial Strategy (BEIS).

We focused on identification of the impact of **smart meters** on customers' energy consumption.

However, the methodology we recommend may also be useful for analyses of other products, services, or interventions that energy suppliers offer customers **where customers' energy consumption is an outcome of interest**.



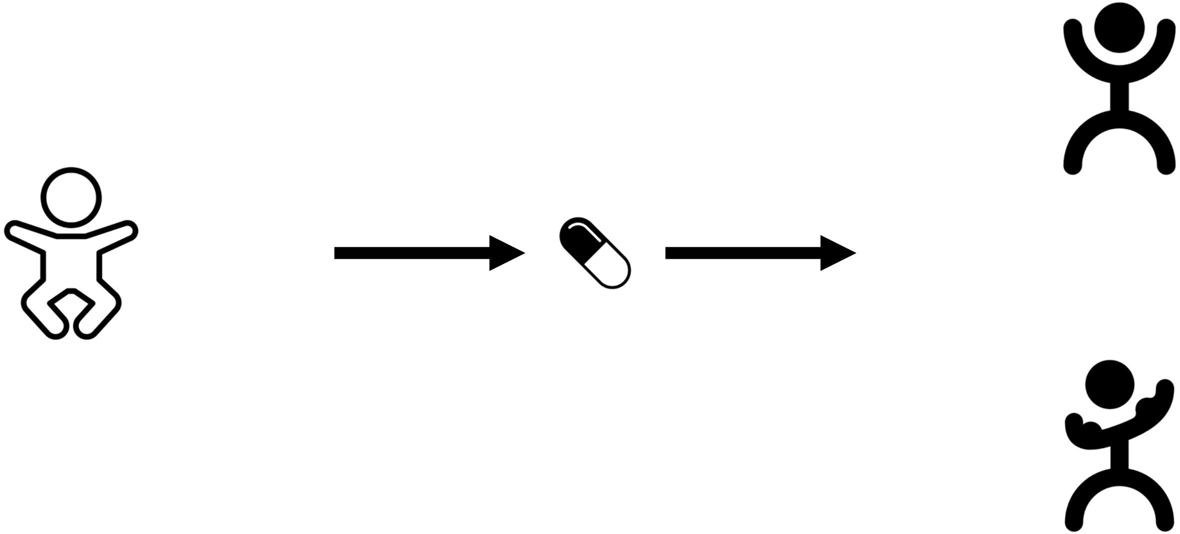
“The critical step in any causal analysis is estimating the counterfactual.”

Hal Varian, Chief Economist, Google

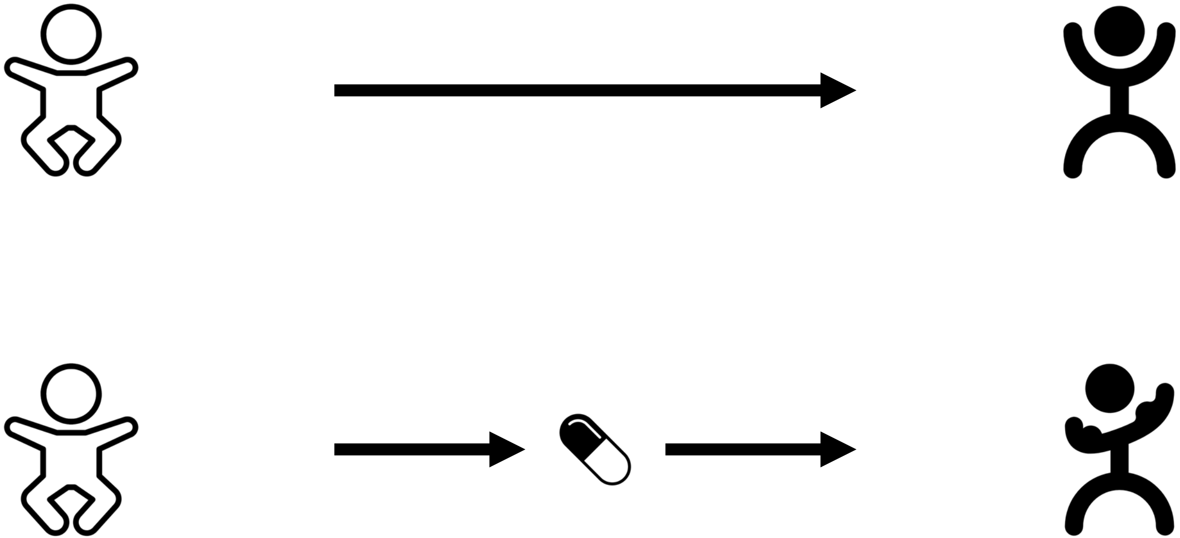
First, a tortured metaphor

THE
BEHAVIOURAL
INSIGHTS
TEAM

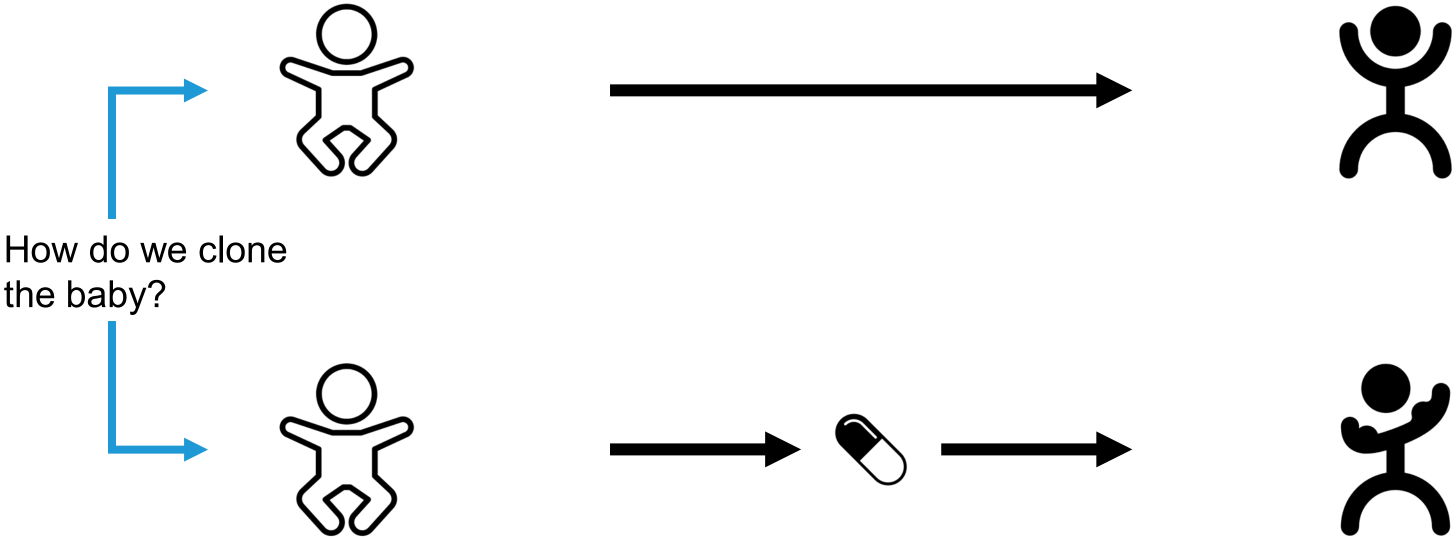
Rigorous analysis is all about the counterfactual



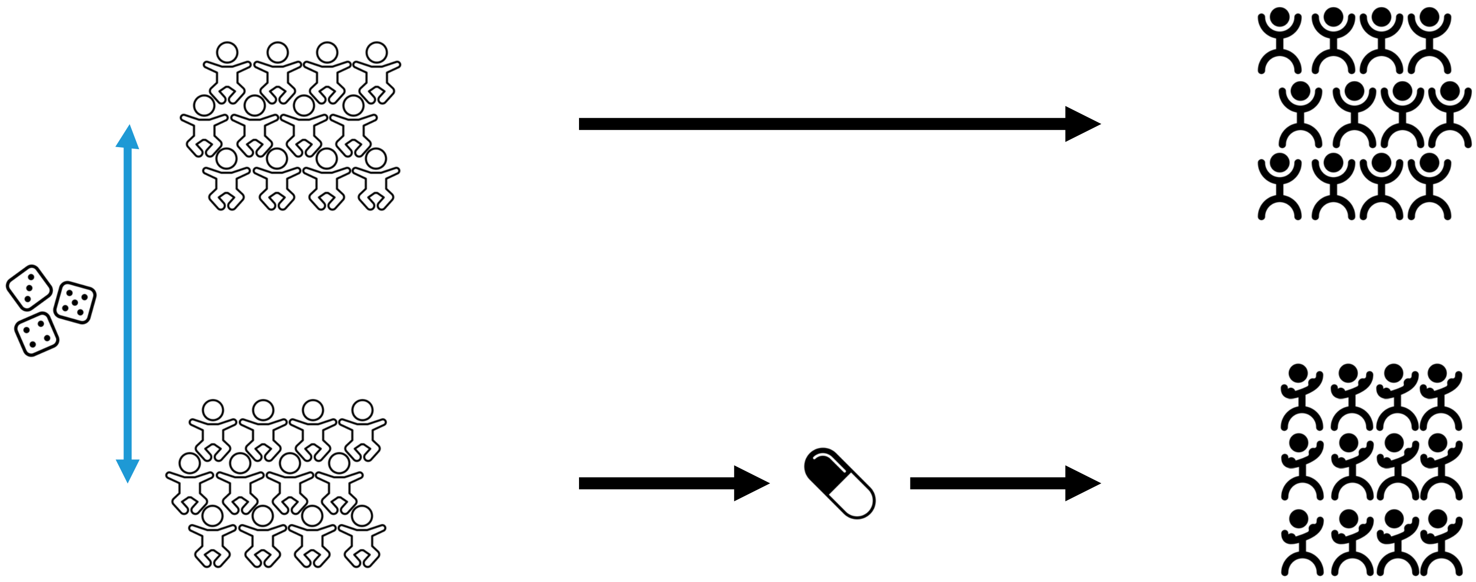
Rigorous analysis is all about the counterfactual



Rigorous analysis is all about the counterfactual



Randomisation is one way of cloning the baby

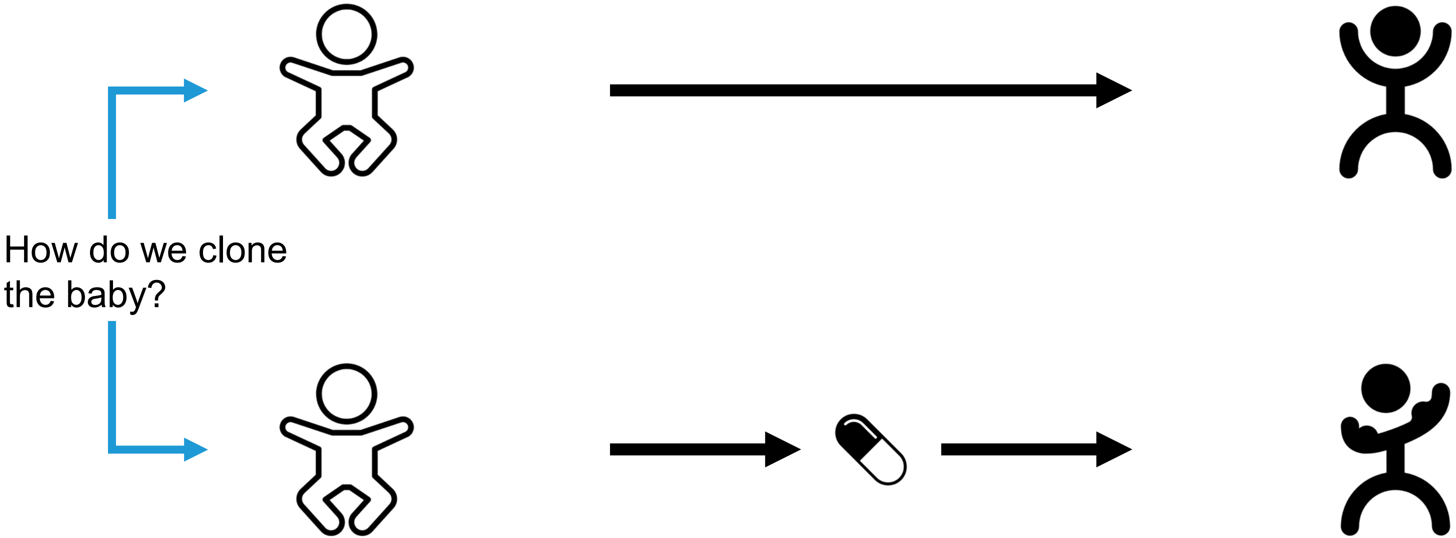


This works because we know **who got the intervention is unrelated to the outcome**

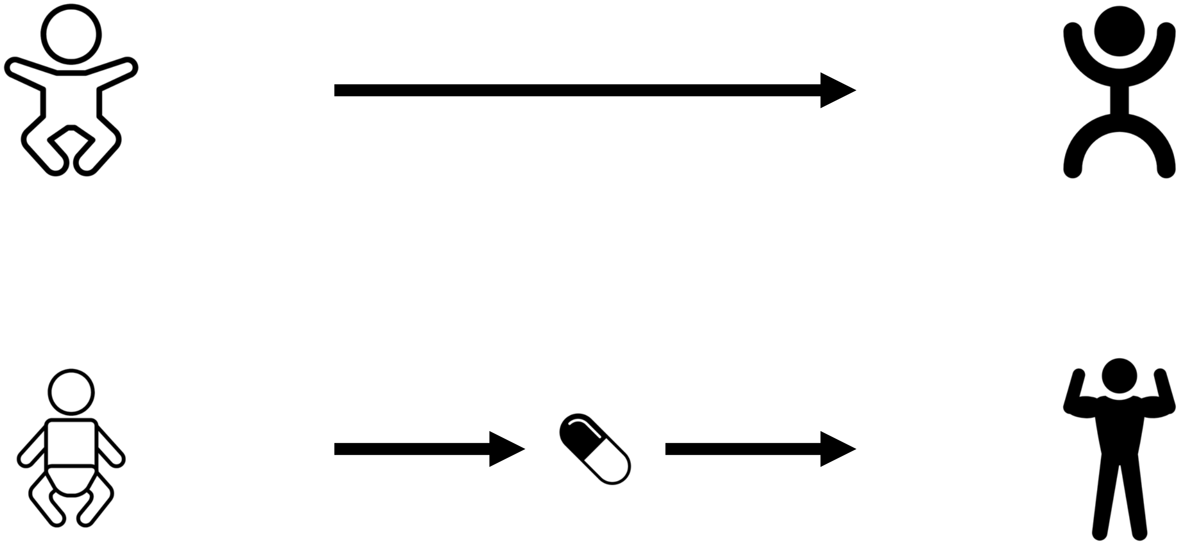
Non-experimental evaluation methods

THE BEHAVIOURAL INSIGHTS TEAM

Rigorous analysis is all about the counterfactual



If randomisation is not possible, you could do your best to find a similar baby



This works because we are **directly constructing a plausible counterfactual**

Experimental and non-experimental evaluations

An example from BIT’s work

THE BEHAVIOURAL INSIGHTS TEAM

Background

Space heating accounts for around 72% of households' gas use.

Traditional heating controls are often unintuitive, and we tend not to use them very efficiently.

Evaluating the Nest Learning Thermostat (NLT)

Can smart heating controls achieve what human behaviour cannot?

In 2014 BIT brought in as independent evaluators to determine energy savings in the UK under ‘real world’ conditions

Four studies concluded between 2014 and 2017. We’ll focus on the first two in this presentation.

Study 1 (2014/15)

- Non-experimental analysis, comparing 2,248 Nest owners to a matched group of 2,248 homes without Nest, using various heating controls
- Indicates **5.8%** (±3.2%), savings on annual household gas consumption ($p<0.001$)
- Suggests savings of 7.8% (±4.3%) on heating system gas use

Study 1 left some key questions unanswered

In particular, we worried about the **risk of selection bias**.

For example, it is possible that households which have purchased a Nest would also be motivated to make other contemporaneous energy-saving changes in their behaviour or their homes, confounding the savings estimate.

Also: The Government was interested in comparing Nests to a control group with a ‘full suite’ of modern heating controls, whereas Study 1 compared Nest households to an unknown aggregate of heating technologies in the control group.

Study 2 (2016/17)

- RCT design avoids selection bias concerns
- Smaller sample (276 homes)
- Standardised benchmark (‘full suite’ of heating controls – a programmer/timer, room thermostat and TRVs)

THE
BEHAVIOURAL
INSIGHTS TEAM.

Study 2 shows similar effect sizes as Study 1

- Estimated **5.6%** (±4.5%) ($p<0.01$) household gas savings across the winter heating period (Oct-April)
- Estimated **4.5 - 5%** (±5.6%) household gas savings across the year
- Estimated **6.8%** saving on heating system gas use
- No loss of thermal comfort (possible improvement)

One key (methodological) takeaway

Studies 1 and 2 involved subtly different comparisons

- Study 1 compared Nest households to an unknown aggregate of heating technologies in the control group.
- Study 2 compared Nest households to households who had a ‘full modern suite’ of technology, defined as a programmable timer, a wall thermostat, and thermostatic radiator valves.
- This means we would expect a lower treatment effect from the Nest in Study 2 compared to Study 1.

With that said, the similarities between the Studies’ results **increased our confidence that matching is an appropriate method to analyse interventions to decrease energy consumption**.

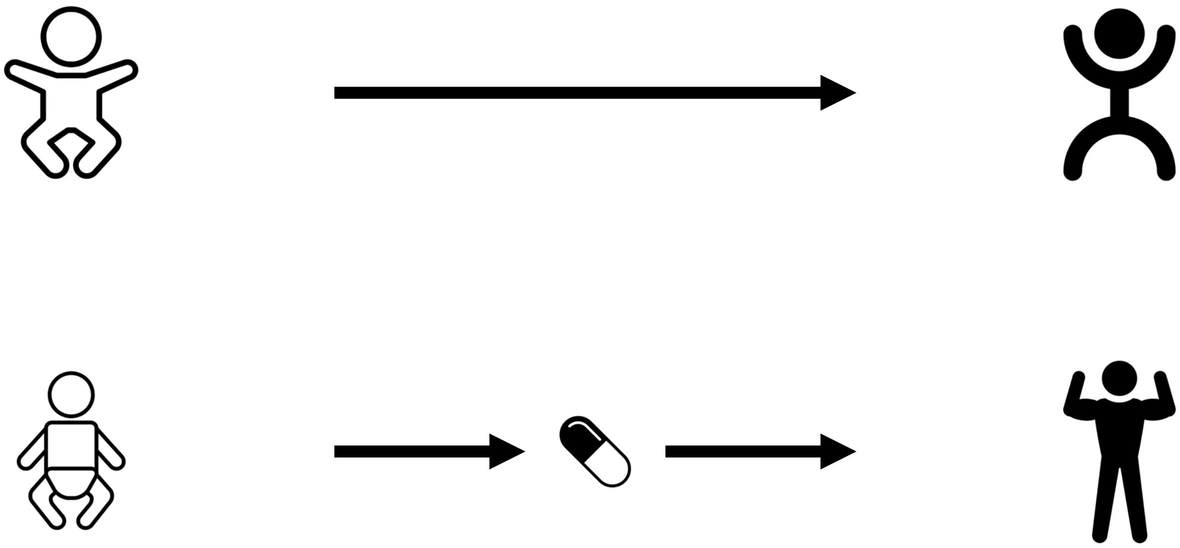
Identifying the impact of smart meters on customers’ energy consumption

Our recommended methodology

THE BEHAVIOURAL INSIGHTS TEAM

The problem: we need to use non-experimental techniques

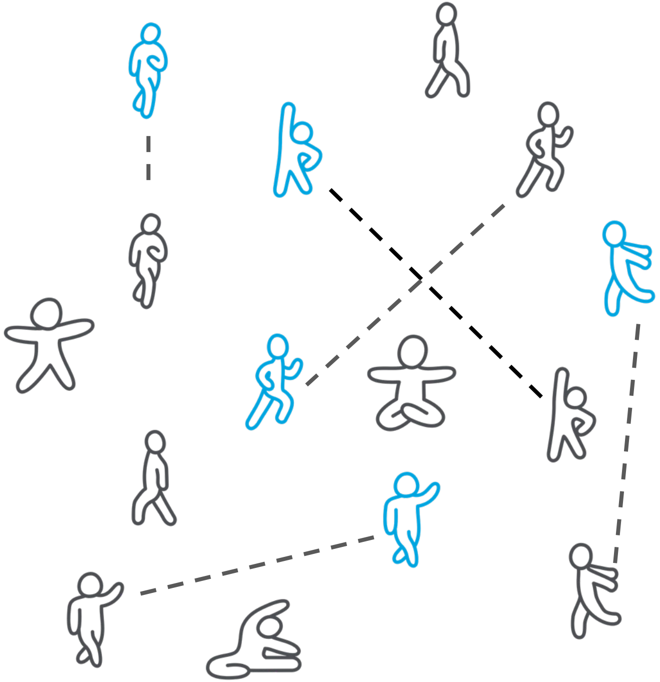
We need to find a **plausible counterfactual**.



Our recommended approach: Matched difference in differences

1. Compare smart-metered customers' consumption to a **matched sample** of traditional-metered customers who have similar characteristics.

Matching makes sense when evaluators have rich data that predicts an outcome well, they have a large universe of potential matches, and participants are not strongly motivated to be in the treatment group. We believe this describes the smart meter context well.



Our recommended approach: Matched difference in differences

1. Compare smart-metered customers' consumption to a **matched sample** of traditional-metered customers who have similar characteristics.

2. Compare the percent change in a household's in consumption/year, rather than raw post-installation consumption – in other words, calculate the **difference in differences**. This is further assurance that differences in baseline consumption do not confound identification of the impact of smart meters.

Note: *In theory, you only need one of the two, but we think combining them provides extra rigour.*

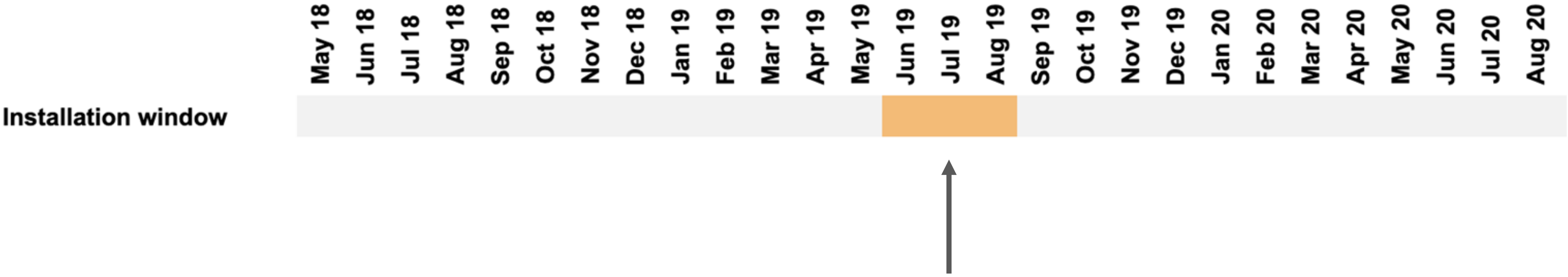
Choosing the smart-metered households and ‘building’ the comparison group

What variables to match on

1. Previous energy consumption/year
2. Region

Note that a smart-metered household may be matched to different traditional-metered households for a supplier's electricity and gas analyses.

Defining the 'installation window'



The window defines the group of smart metered households

1. Matching on pre-installation consumption/year

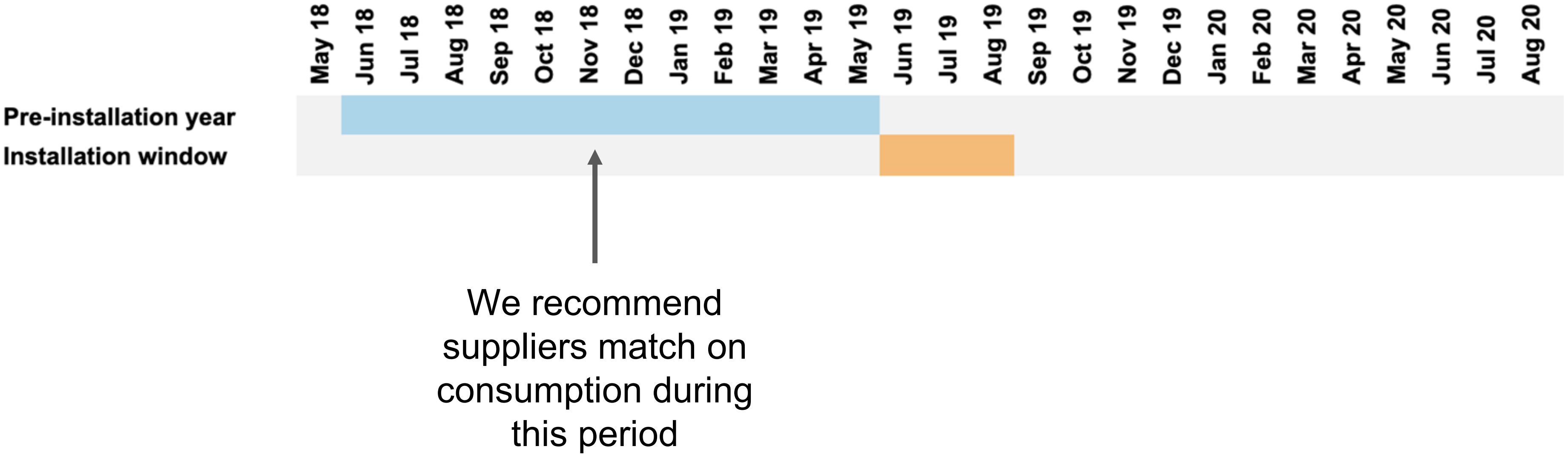
We recommend prioritising matches on pre-installation consumption/year, using tight matches if possible: **±50 kWh/year bands for electricity** and **±200 kWh/year bands for gas**.

Widen these bands if sample size would otherwise be insufficient. We recommend using a maximum band size of ±200 kWh/year for electricity and ±800 kWh/year for gas.

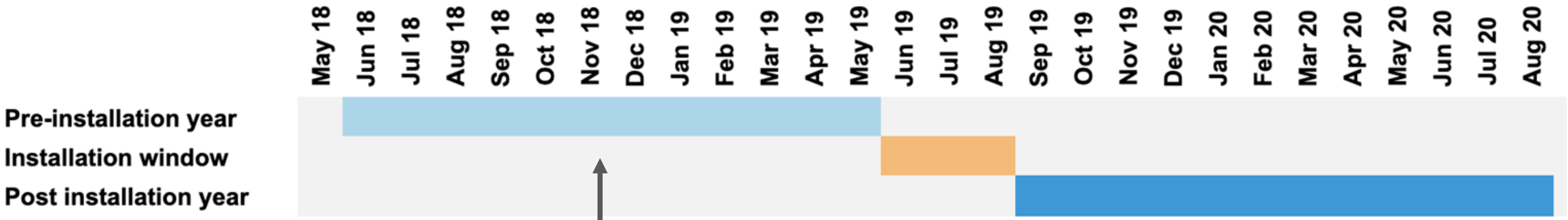
Examining 3% reductions in consumption/year, by Typical Domestic Consumption Value

	Ofgem TDCV category	Consumption/ year (kWh/year)	Consumption/ year with 2% decrease (kWh/year)	Difference (kWh/year)
Gas	Low	8,000	7,760	240
	Medium	12,000	11,640	360
	High	17,000	16,490	510
Electricity (profile class 1)	Low	1,900	1,843	57
	Medium	3,100	3,007	93
	High	4,600	4,462	138

Defining the 'pre-installation year'



Defining the ‘post-installation year’



We recommend suppliers match on consumption during this period

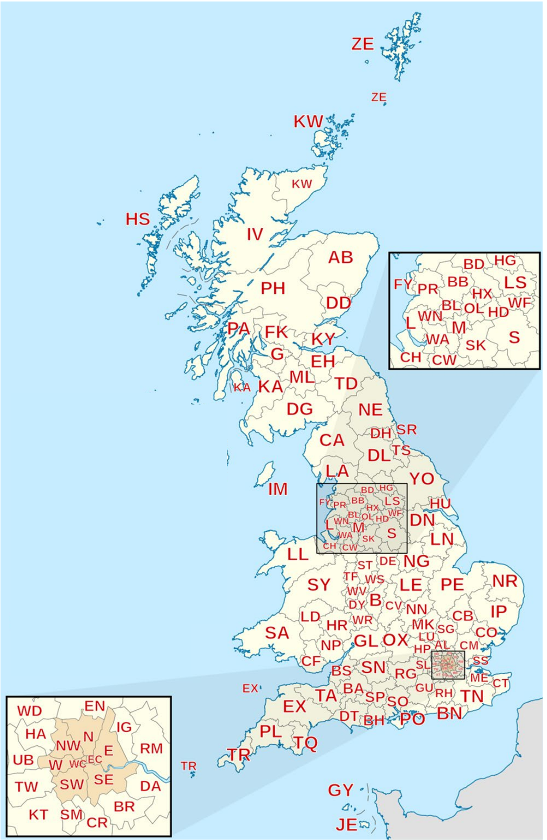
And compare consumption during this period

2. Matching on region

Region-specific events and trends – such as weather – affect energy consumption in that region.

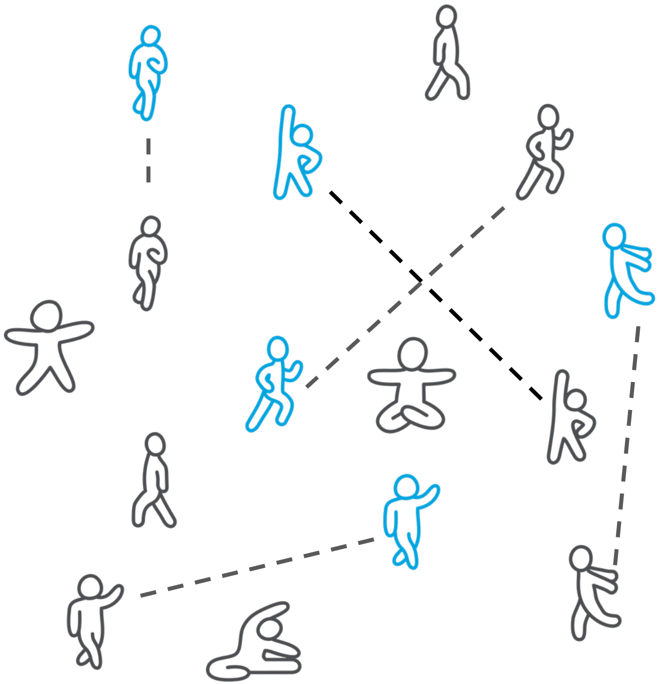
Options for region matching include:

- PES region
- Outer postcode



A note on matching: one-to-one matching is good

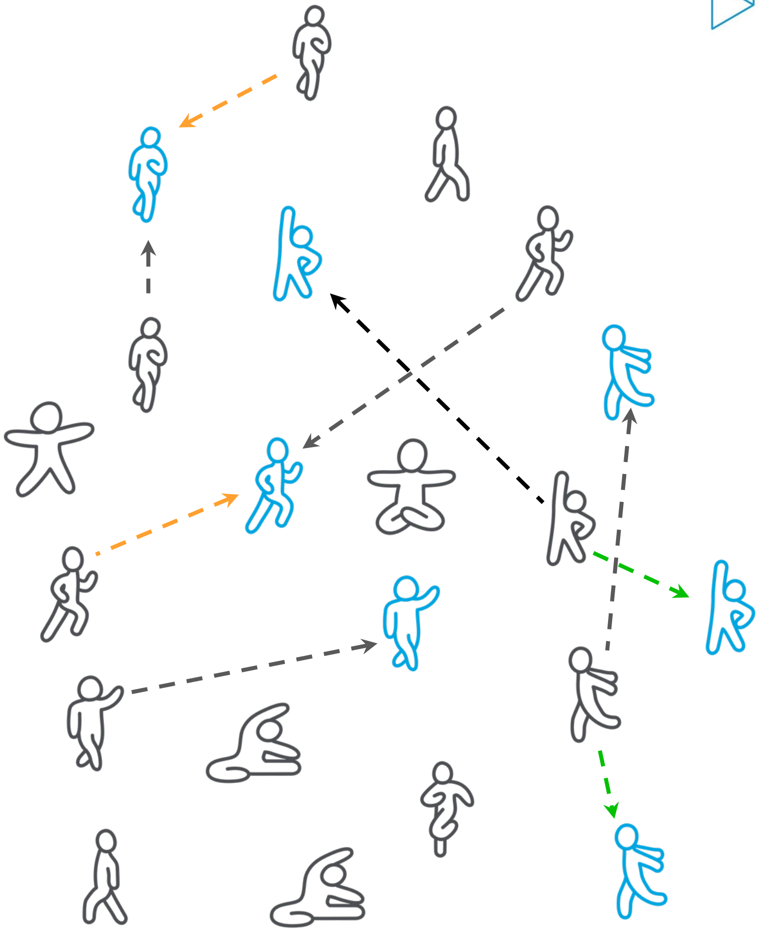
One-to-one matching finds a single match for each unit in the ‘treatment’ group.



But many-to-many matching is better!

Many-to-many matching allows more precise analysis with the same sample size by making more efficient use of the available data.

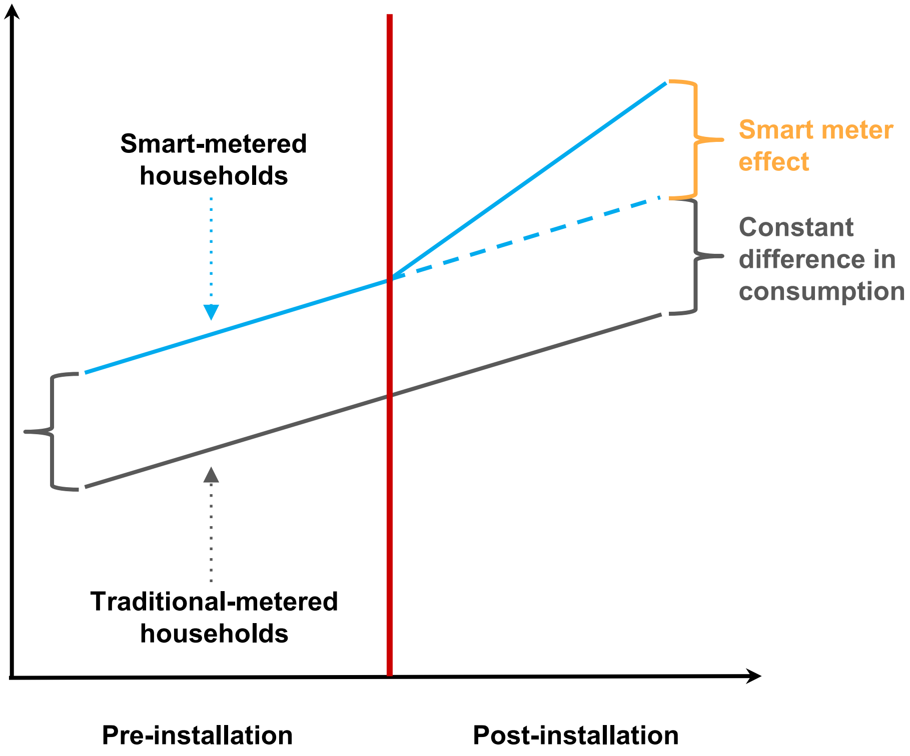
- **Multiple traditional-metered households can serve as comparisons for a smart-metered household.** This means estimates make use of these extra comparisons (rather than randomly choosing just one traditional-metered household to serve as the match).
- A traditional-metered household can serve as the comparison household for multiple smart-metered households. **This means fewer households are unmatched.**



Comparing consumption in smart-metered households and their matched sample

Difference in differences

The differences in differences technique calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome for the treatment group, compared to the average change over time for the control group.



Calculating a household's percent change in energy consumption

For each household *i*, calculate:

$$D_i = (A_i - B_i) / B_i$$

where:

- D_i is the difference in consumption post-installation versus pre-installation, scaled by pre-installation consumption
- A_i is the post-installation consumption for the household (in kWh/year)
- B_i is the pre-installation consumption for the household (in kWh/year)

So, in summary…

1. Compare smart-metered customers' consumption to a **matched sample** of traditional-metered customers who have similar characteristics.

2. Compare the percent change in a household's in consumption/year, rather than raw post-installation consumption – in other words, calculate the **difference in differences**. This is further assurance that differences in baseline consumption do not confound identification of the impact of smart meters.

Limitations to validity

Limitations to external validity

- Exclusion of **frequent switchers** and **customers who rarely or never give meter readings** – these excluded customers may respond differently to obtaining a smart meter than those in the analyses.
- The analysis results apply to the smart-metered customers in the **installation window** suppliers analyse – insofar as these customers or the installation window were atypical, their experience might not generalise to other customers' experience.

Limitations to internal validity

- This methodology **excludes traditional-metered customers who do not give meter readings** during the year in which their consumption is compared to smart-metered customers – but there is no such filter for smart-metered customers.
    - *However, we expect this bias to be modest – and it may point in the opposite direction to the threat to internal validity mentioned immediately above.*
- Matching does not guarantee groups that are balanced, on average, between unobservable characteristics – for example, **they may be unbalanced on enthusiasm to change their behaviour**.
    - *However, we believe that a quality matching approach mitigates the risk of unobserved differences between the two groups confounding identification of smart meters' impact.*

More on internal validity: is matching sufficient?

Economists can be skeptical about matching as an effective way to uncover the treatment effect of a non-randomised intervention. David McKenzie recently discussed this issue on the World Bank blog. He also helpfully discussed rules of thumb for when researchers **should** consider matching.

To make a long story short, we believe smart meter uptake meets some of McKenzie's rules of thumb:

- Smart meter installation is driven by supplier outreach, more than active customer decisions or requests. Indeed, smart meter installation is characterized by supplier bottlenecks that mean customer selection is quasi-random – though suppliers themselves will have insights into their own processes that may inform this logic.

For more details, see our full reports

https://www.bi.team/blogs/new-guidance-on-conducting-energy-consumption-analysis/

https://www.bi.team/publications/evaluating-the-nest-learning-thermostat/

Get in touch: andrew.schein@bi.team