Counterfactual benefits and challenges: why and how?

Rapid evaluation in health care 2020

January 2020

julia.shen1@lshtm.ac.uk

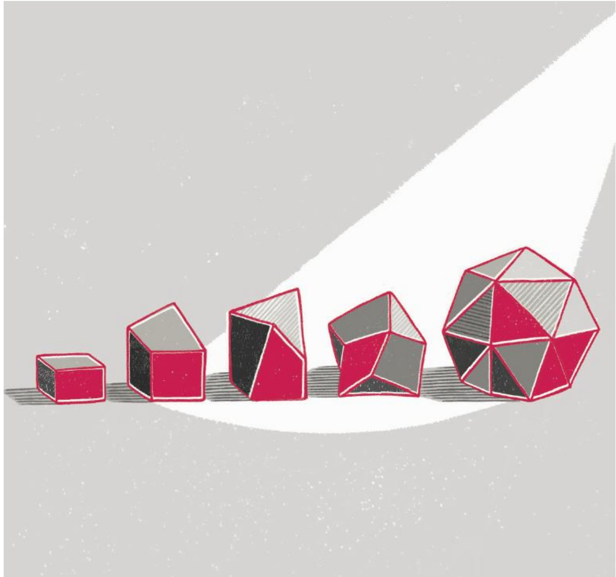
The Health Foundation

NHS ENGLAND
NHS IMPROVEMENT

About us

The Improvement Analytics Unit (IAU) is a unique partnership between NHS England and the Health Foundation that evaluates complex local initiatives in health care in order to support learning and improvement.

www.health.org.uk/IAU

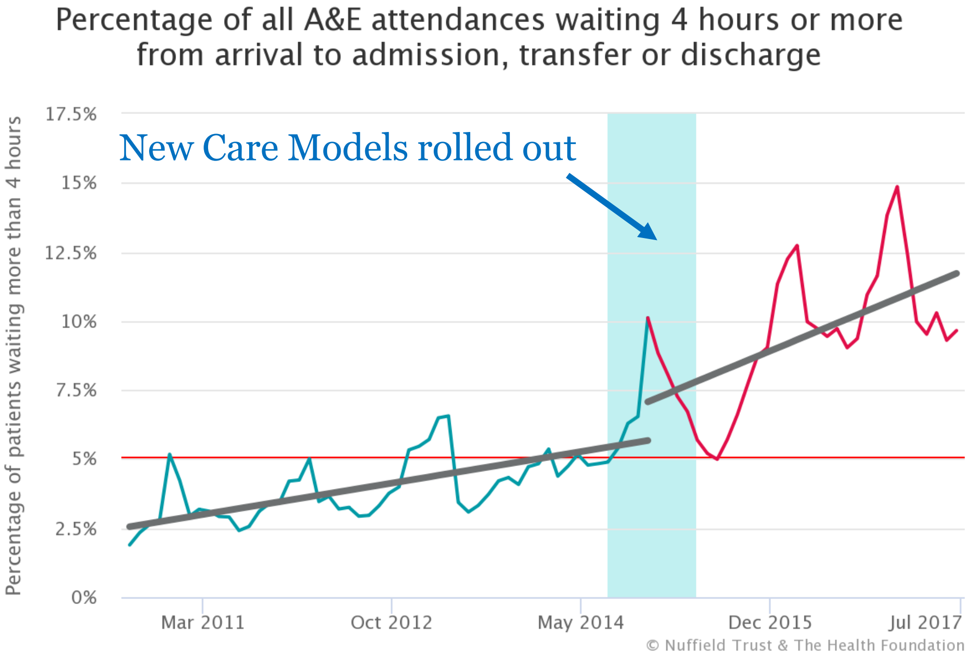


We shine a light on how to make successful change happen

Why do we care about counterfactuals?

Solving emergent challenges in complex health systems is difficult…

Should we conclude that the New Care Models are ***causing*** more delayed transfers of care? Probably not!



Why do we care about counterfactuals?

… 'common sense' solutions are not obviously common knowledge…

Research article | Open Access | Open Peer Review | Published: 10 May 2018

The effects of integrated care: a systematic review of UK and international evidence

Susan Baxter, Maxine Johnson, Duncan Chambers, Anthea Sutton, Elizabeth Goy

BMC Health Services Research **18**, Article number: 350 (2018) | Cite this article

Why do we care about counterfactuals?

… 'common sense' solutions are not obviously common knowledge…

Results

One hundred sixty seven studies were eligible for inclusion. Analysis indicated evidence of perceived improved quality of care, evidence of increased patient satisfaction, and evidence of improved access to care. Evidence was rated as either inconsistent or limited regarding all other outcomes reported, including system-wide impacts on primary care, secondary care, and health care costs. There were limited differences between outcomes reported by UK and international studies, and overall the literature had a limited consideration of effects on service users.

Conclusions

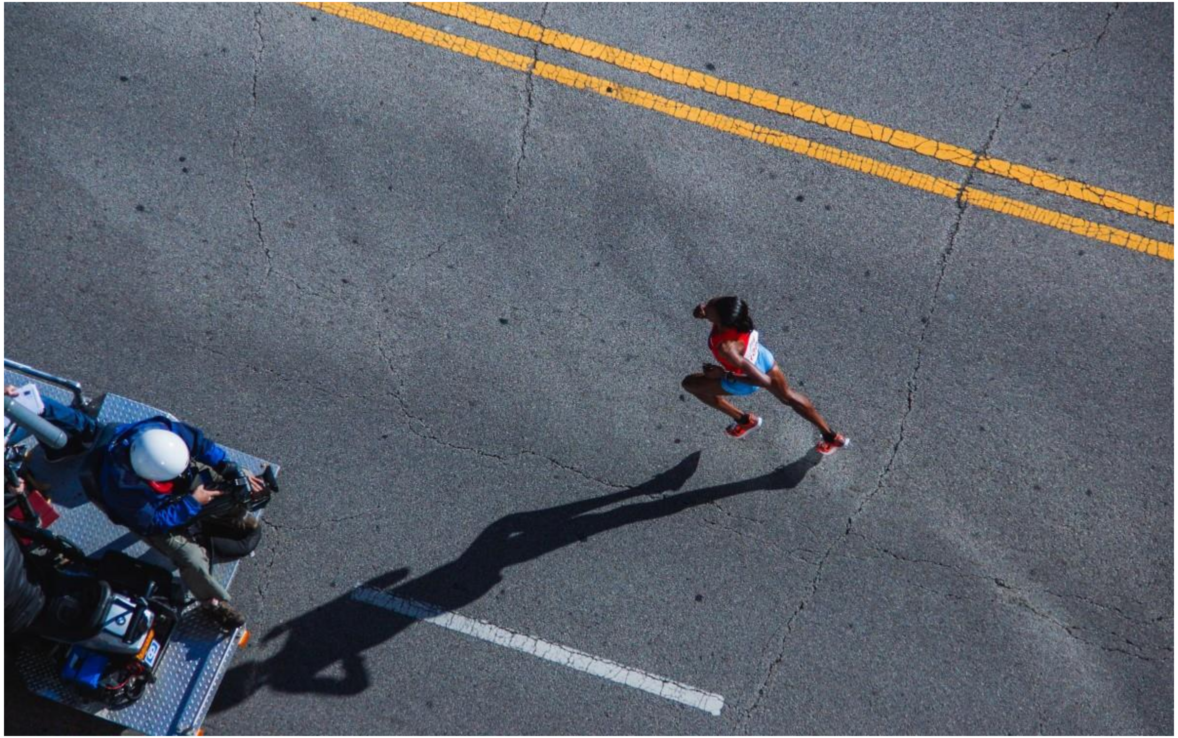
Models of integrated care may enhance patient satisfaction, increase perceived quality of care, and enable access to services, although the evidence for other outcomes including service costs remains unclear. Indications of improved access may have important implications for services struggling to cope with increasing demand.

Trial registration

Prospero registration number: 42016037725.

Why do we care about counterfactuals?

… and much of the data we access imperfectly captures reality



SUS and other NHS data, like all administrative data, capture a '**shadow**' of the reality we actually care about.

Why do we care about counterfactuals?

… and much of the data we access imperfectly captures reality

BMJ Quality & Safety

Latest content Curre

Home / Archive / Volume 24, Issue 8

Article Text

Article info

Citation Tools

Share

Original research

Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data

Sarah R Deeny, Adam Steventon

Author affiliations +

Abstract

Socrates described a group of people chained up inside a cave, who mistook shadows of objects on a wall for rea
comes to mind when considering 'routinely collected data'—the massive data sets, generated as part of the rou
the modern healthcare service. There is keen interest in routine data and the seemingly comprehensive view of
offer, and we outline a number of examples in which they were used successfully, including the Birmingham Ow

Defining causality: a bit of algebra

Neyman-Rubin causal model (‘potential outcomes framework’)

Y = outcome Y=1 (yes) and Y=0 (no)

X = intervention X=1 (treated) and X=0 (not treated)

Observations averaged over many individual units or times ***t***

Estimating causal effects (typical ‘estimands’ in healthcare)

$\Pr(Y_t = 1 \mid X_t = 1) - \Pr(Y_t = 1 \mid X_t = 0)$ *risk difference*

$\Pr(Y_t = 1 \mid X_t = 1) \,/\, \Pr(Y_t = 1 \mid X_t = 0)$ *risk ratio*

Technical articles repeat this notation > the maths looks harder than the concepts actually are to understand and apply.

Fundamental problem: *Within the same individual or unit, we only ever observe one outcome (Y=1 or Y=0), and usually we can only treat them or not (X=1 or X=0).* **The counterfactual is the imagined opposite of what we saw.**

Defining causality

Judea Pearl's "ladder"

3. Imagination: Watering makes plants grow, and I can do this with a hose when it doesn't rain.

$$\Pr(Y = 1 \mid do(X) = 0, do(X^*) = 1)$$

2. Intervention: If I cover my crops when it rains, they will not grow.

$$\Pr(Y = 1 \mid do(X) = 1)$$
$$\Pr(Y = 0 \mid do(X) = 0)$$

1. Association: Crops seem to grow better after rain.

$$\Pr(Y = 1 \mid X = 1) = \frac{\Pr(Y = 1 \cap X = 1)}{\Pr(X = 1)}$$

Defining causality

Climbing the ladder

Counterfactual study designs are necessary to actually test hypotheses and ultimately to ensure we are making better policy decisions. Most commonly, these are **RCTs**.

If we must use observational data for evaluation, we ought to use **quasi-experimental techniques** (e.g. matched controls, synthetic controls, instrumental variables, regression discontinuity) and see **replicated results.**

Most observational studies, even if conducted carefully, only establish that we **might** have a causal relationship. **Pre-post and ecological studies** are useful to generate hypotheses, but are **insufficient** on their own find ‘true’ effects of interventions and conclude that a policy universally ‘works’.

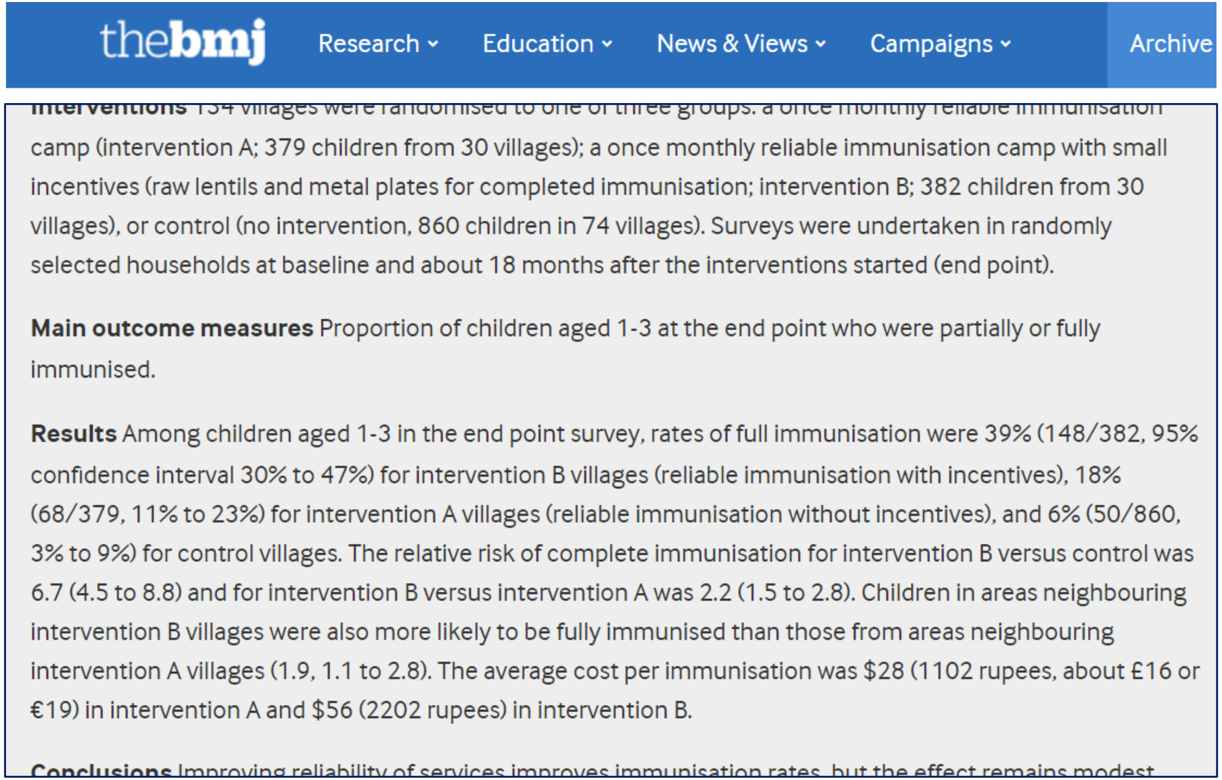
Randomised controlled trials in policy

Careful application of causal experiments to evaluate international development interventions won the 2019 Nobel Prize in Economics for applied researchers at MIT and Harvard.

ipa INNOVATIONS FOR POVERTY ACTION

“ Since its founding, [Innovations for Poverty Action]’s infrastructure for carrying out field experiments helped enable a proliferation of rigorous evaluations. IPA's largest office, in Kenya, was started in 2005 to carry out projects that [Abhijit Banerjee, Esther Duflo, Michael Kremer] and other researchers had started. As of 2019, IPA has implemented over 830 evaluations of programs in 51 countries, each led by leading researchers, such as the three Laureates, to test key questions about how to alleviate poverty. ”

Randomised controlled trials in policy



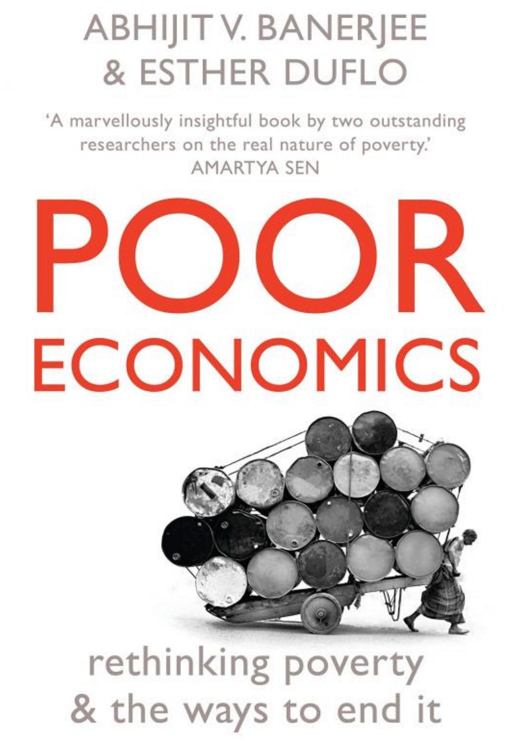
thebmj Research Education News & Views Campaigns Archive

Interventions 134 villages were randomised to one of three groups: a once monthly reliable immunisation camp (intervention A; 379 children from 30 villages); a once monthly reliable immunisation camp with small incentives (raw lentils and metal plates for completed immunisation; intervention B; 382 children from 30 villages), or control (no intervention, 860 children in 74 villages). Surveys were undertaken in randomly selected households at baseline and about 18 months after the interventions started (end point).

Main outcome measures Proportion of children aged 1-3 at the end point who were partially or fully immunised.

Results Among children aged 1-3 in the end point survey, rates of full immunisation were 39% (148/382, 95% confidence interval 30% to 47%) for intervention B villages (reliable immunisation with incentives), 18% (68/379, 11% to 23%) for intervention A villages (reliable immunisation without incentives), and 6% (50/860, 3% to 9%) for control villages. The relative risk of complete immunisation for intervention B versus control was 6.7 (4.5 to 8.8) and for intervention B versus intervention A was 2.2 (1.5 to 2.8). Children in areas neighbouring intervention B villages were also more likely to be fully immunised than those from areas neighbouring intervention A villages (1.9, 1.1 to 2.8). The average cost per immunisation was $28 (1102 rupees, about £16 or €19) in intervention A and $56 (2202 rupees) in intervention B.

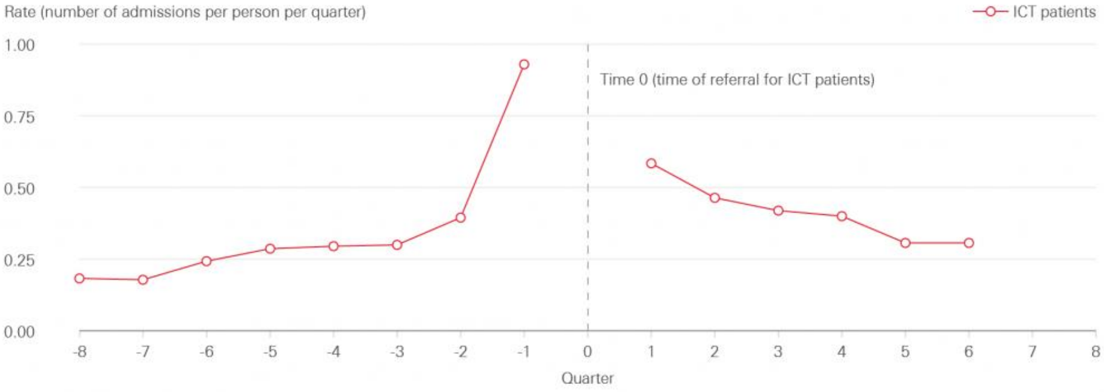
Conclusions Improving reliability of services improves immunisation rates, but the effect remains modest



Aim of counterfactual study designs: **pragmatically replicate RCT estimates**

1. Compare like-with-like in terms of **time** → *avoid regression to the mean*

Figure 1. Rates of emergency admissions for ICT patients: comparison before and after referral



The Health Foundation © 2018

Source: Lloyd, T. et al. The impact of integrated care teams on hospital use in North East Hampshire and Farnham. 2018.

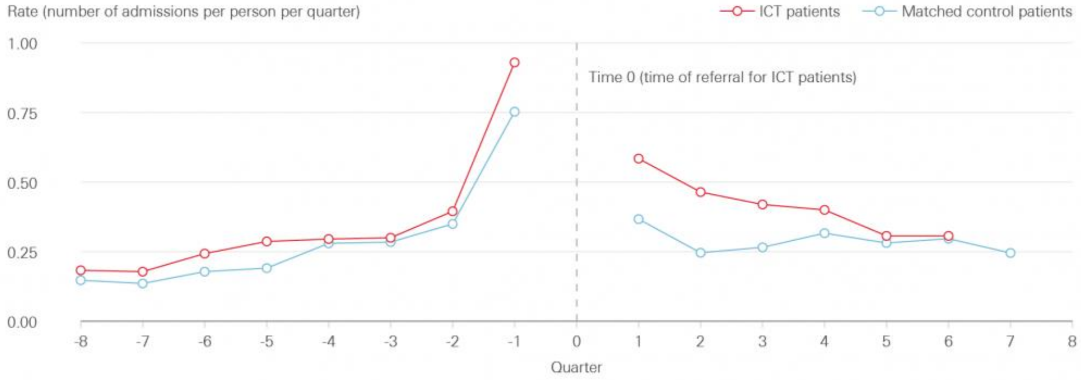
At first glance, this intervention in North East Hampshire and Farnham looks positive and effective for reducing emergency admissions…

Aim of counterfactual study designs: **pragmatically replicate RCT estimates**

1. Compare like-with-like in terms of **time** → *avoid regression to the mean*

… but we discover that observed effect is misleading and may even be the opposite of reality when we have a robust counterfactual of comparable patients who did not receive the intervention over the same time.

Figure 3. Rates of emergency admissions: comparison of ICT and matched control patients before and after referral



The Health Foundation © 2018

Source: Lloyd, T. et al. The impact of integrated care teams on hospital use in North East Hampshire and Farnham. 2018.

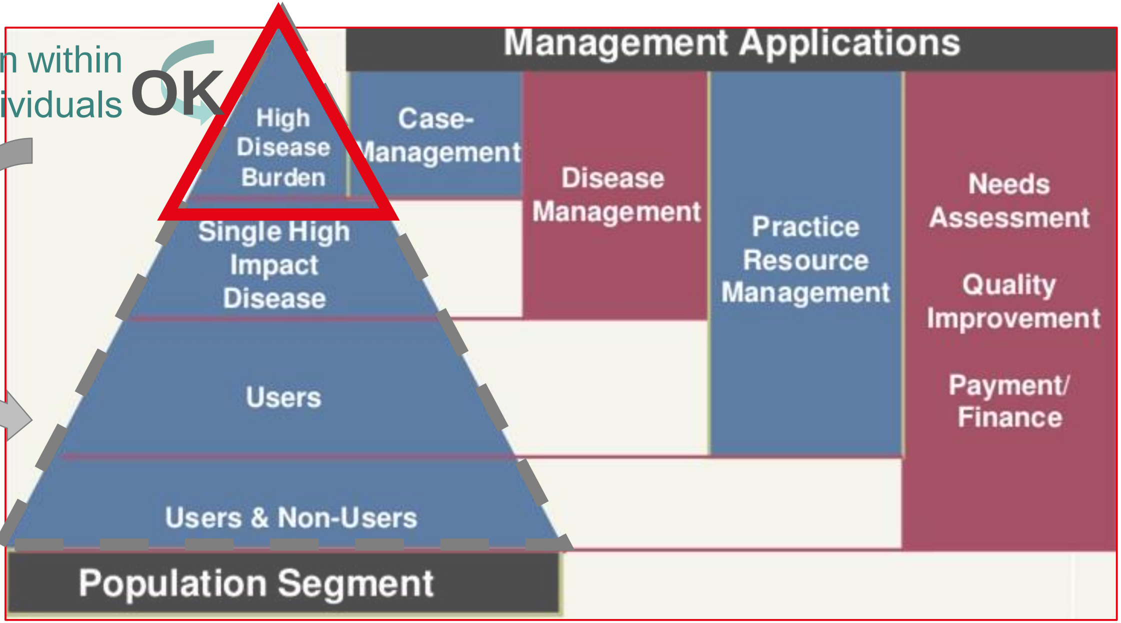
Aim of counterfactual study designs: **pragmatically replicate RCT estimates**

2. Compare like-with-like in terms of **characteristics** → *avoid confounding*

like-for-like comparison within 'exchangeable' individuals OK

confounded comparison due to fundamentally different groups

In other cases, such as targeted interventions based on risk stratification, improper counterfactual specification might give us null or negative results, as when unfairly comparing to a healthier population, for an otherwise effective policy.



Two examples of counterfactual methods today

Synthetic controls (Stefano Conti)

using time series data from a weighted group of comparable systems or areas

Genetic matching (Paul Seamer)

using observed baseline covariate and outcome data from a selected group of comparable control individuals

Briefing

December 2017

Briefing: The impact of redesigning urgent and emergency care in Northumberland

Home > Publications > Evaluation of an Integrated Mental Health Liaison Service (Rapid Assessment Interface and Discharge Service) in Northern Irelan

Evaluation of an Integrated Mental Health Liaison Service (Rapid Assessment Interface and Discharge Service) in Northern Ireland

MENTAL HEALTH SERVICES

Why don’t we do more counterfactual research?

Good reasons

1) Problem and solution are actually well understood –cf. satirical *BMJ* protocol for an RCT of parachutes when jumping from airplanes

2) Unethical experimentation, e.g. organ transplantation cannot be randomized

Solvable reasons

3) Lack of sufficient-quality data → investment in data infrastructure + capacity

4) Lack of time or money → evaluation is much faster when #3 is sorted and pays off in the long run… we ought to prefer durable evidence from one careful study over repetition of compromise studies that leave us uncertain

5) Lack of confidence in applying methods → resources and advice freely available, including from us at the **Improvement Analytics Unit**

Accessible resources

1. Cunningham S. *Causal inference: The mixtape*. v1.7, 2020. Available free online at http://scunning.com/cunningham_mixtape.pdf.
2. Hernan M and Robins J, *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC, 2020. Available free online at https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
3. Lloyd T. "Why before-and-after analyses can give misleading results", Sept 2018, at https://www.health.org.uk/newsletter-feature/why-before-and-after-analyses-can-give-misleading-results.
4. Pearl J and Mackenzie D, *The Book of Why: The new science of cause and effect.* New York: Basic Books, May 2018.
5. Stock M. "Climbing the ladder of causality", June 2018. See the summary overview of Pearl's reasoning in the above book at https://michielstock.github.io/causality/.

[for examples in previous pages, please click on images for links]

Thank you

The Health Foundation

NHS

NHS ENGLAND
NHS IMPROVEMENT