THE POLICY INSTITUTE

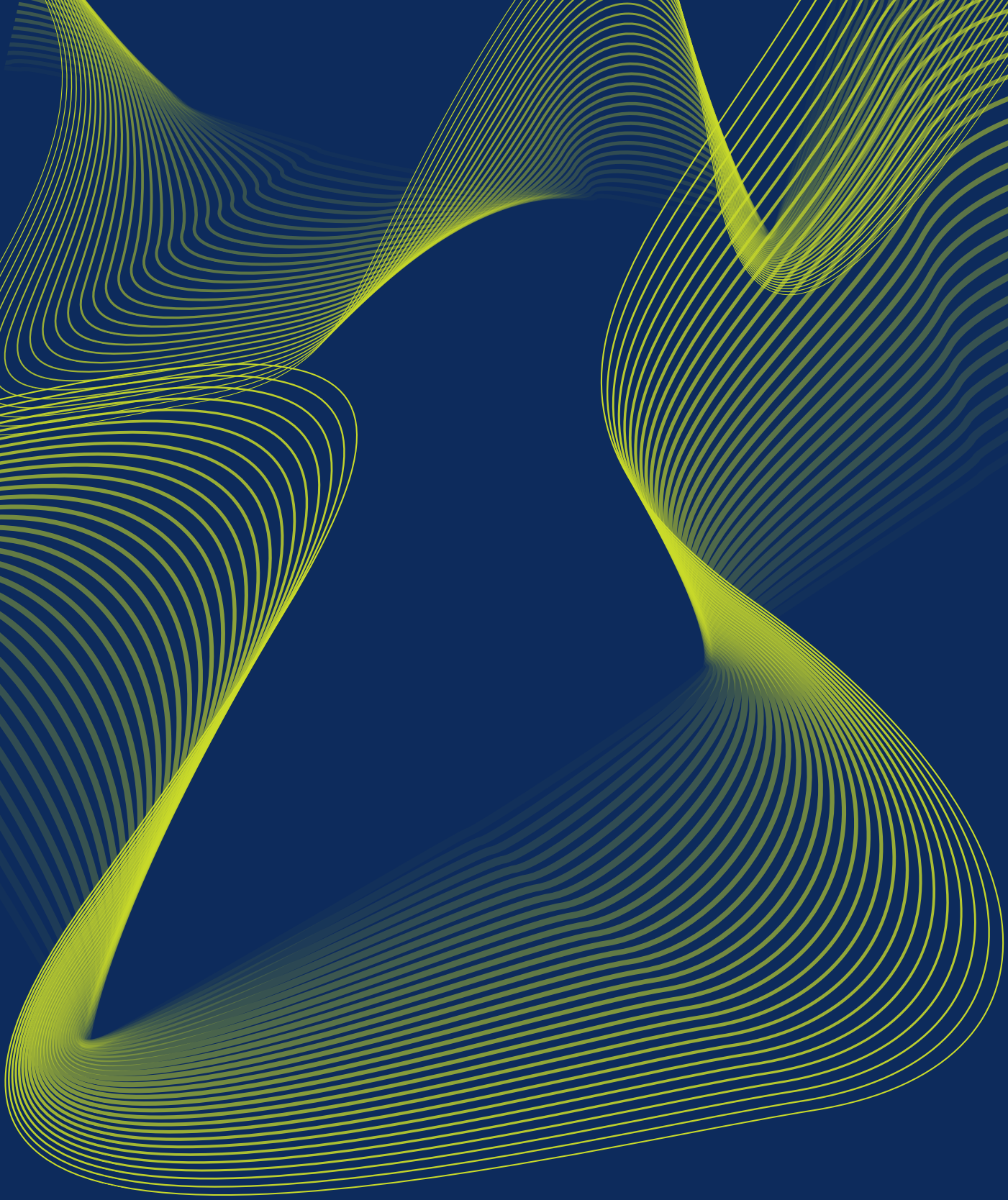
KING'S College LONDON

What Works, faster

Towards a "rapid method"

Michael Sanders and Ella Whelan

February 2022



Introduction

Governments around the world have increasingly recognised the value in using either randomised controlled trials (RCTs) or quasi-experimental approaches to evaluate public policy.

Unlike other forms of research and evaluation, these methods, often bundled under the title of "What Works", seek to reliably answer questions roughly in the form "If I do X, what will happen to Y?" for various values of both X and Y.

In RCTs, participants in a study are assigned at random to either receive a new intervention (the treatment group) or business as usual (the control group), and their outcomes are compared. If there are enough units to be randomised, the two groups will be statistically identical in the absence of the intervention, and so any differences that are observed between the two can be said to have been caused by the intervention. Quasi-experimental approaches make use of statistical techniques, and quirks in the way that interventions are delivered or implemented, to approximate the effect of randomisation in its absence.

Since the end of the second world war, medicine has undergone a revolution in which the judgement of clinicians, often based on an abundance of anecdotal experience, has been supplemented, and in many cases replaced, with the widespread use of randomised controlled trials for those interventions where they are possible, and for better use of data for interventions where they are not.

Other fields, and social policy more generally, has been slower to adopt these approaches, in part due to substantial resistance to randomisation across public service professionals outside of medicine[1]. The 21st century has seen, however, a rise in the use of randomisation in other fields, prominently including policing and education.

In a British context, an important step in this process has been the establishment of the What Works Network – a network of centres, starting with the National Institute for Health and Care Excellence (NICE) in 1999, and followed by the Education Endowment Foundation in 2011, which all serve to advance the use of "What Works" methodologies in their respective fields.

At the time of writing, there are 13 centres, covering policy areas from health to housing, and from before a child is born to the end of the life of the oldest adult. Although these centres vary radically in size and activities, they share a core philosophical belief in the possibility and importance of answering questions as we describe above; "If we do X, what will happen to Y?". It is important, however, not to claim victory. First, there is still a great deal of policy – the overwhelming majority – for which we do not have an answer to this question. Moreover, while we have advanced our progress in producing high quality causal evidence, the production of new policy initiatives continues to massively outstrip the production of new evaluations of their effectiveness. The golden age of policy evaluation sees a fall in the proportion of all policies that go un-evaluated, but a rise in the absolute number of policies.

[1] It should be noted here that medicine itself was not an enthusiastic adopter of trials – change was achieved gradually and in the face of substantial resistance.

There is a frustration with the amount of time that evaluation takes. With short political cycles, people would like the results right now. Good research takes time, attention to detail, and diligence. With more funding we could maybe go a little bit faster, but not much. After all, people need to be recruited, consented, randomised, receive the intervention, and then we need to wait long enough for it to have had an effect. Unlike medicine, we do not have standard procedures for recruitment, and evaluators usually do not work in our equivalent hospitals, where patients are in ready supply. Also unlike medicine, where "receiving the intervention" might be as simple as popping a pill in your mouth, swallowing and washing it down with some water, some interventions – like the Family Nurse Partnership – might take years to even administer.

At this point, our instincts are perhaps to grumble. Why can our funders not understand that research takes time? Or that, if you can only pick two from "good, quick, and cheap", and governments always prioritise "cheap", they should really choose "good" as their second option, even if this means results take longer?

Grumbling is of course one option, and one that we have ourselves engaged in from the comfort of our own office (or Zoom calls), or at conferences. But we should also interrogate why the people we work with – often, or even usually, public servants – want us to hurry. Yes, some of it is driven by political expedience, the timing of spending reviews and so on, but some of it – we'd argue most – comes from an altogether more positive place. Every year that goes past without us knowing what works, the problems we are trying to fix remain unfixed; the world remains a poorer, less just place. There are real, human consequences to delays in evaluations. Here, those who are pressing us to be faster are the truest allies to those of us pushing for more evaluation – both believe that without evaluation, the world is worse. We want more evaluation, and they want it now.

This paper considers what we can do better and faster in the pursuit of a system of policymaking and policy evaluation to achieve better outcomes for the people we serve.

The case of the Oxford Astra-Zeneca vaccine.

In 2020, the world was gripped by the Covid-19 pandemic, and researchers around the world began rapidly to try and develop a vaccine against the virus. It's fair to say that the vaccine development world got the drop on the rest of us in terms of taking the virus seriously, with development beginning back in January of 2020, when lockdowns in most of the world were but a twinkle in a newspaper columnist's eye.

In another sense, vaccine scientists were even further ahead of us. They had been on the look out for "virus X" – something highly transmissible and fairly fatal, which could cause a global pandemic of the like we've all been living through – for years, knowing that it was a distinct possibility and that, when it came along, a vaccine would needed and needed much more quickly than the five plus years vaccine development and licensing typically takes. The story of how this happened in the case of the Oxford AstraZeneca vaccine is retold by Professors Sarah Gilbert and Catherine Green, two of the Oxford researchers who developed the vaccine in record time.

As the two professors are at pains to point out, there were no corners cut and no safety compromised in the pursuit of developing the vaccine as quickly as possible. Instead, the Oxford team did everything as fast as they could, prioritising getting work started and then working out how to pay for it later, and were helped eventually by massive investments from governments in vaccine development.

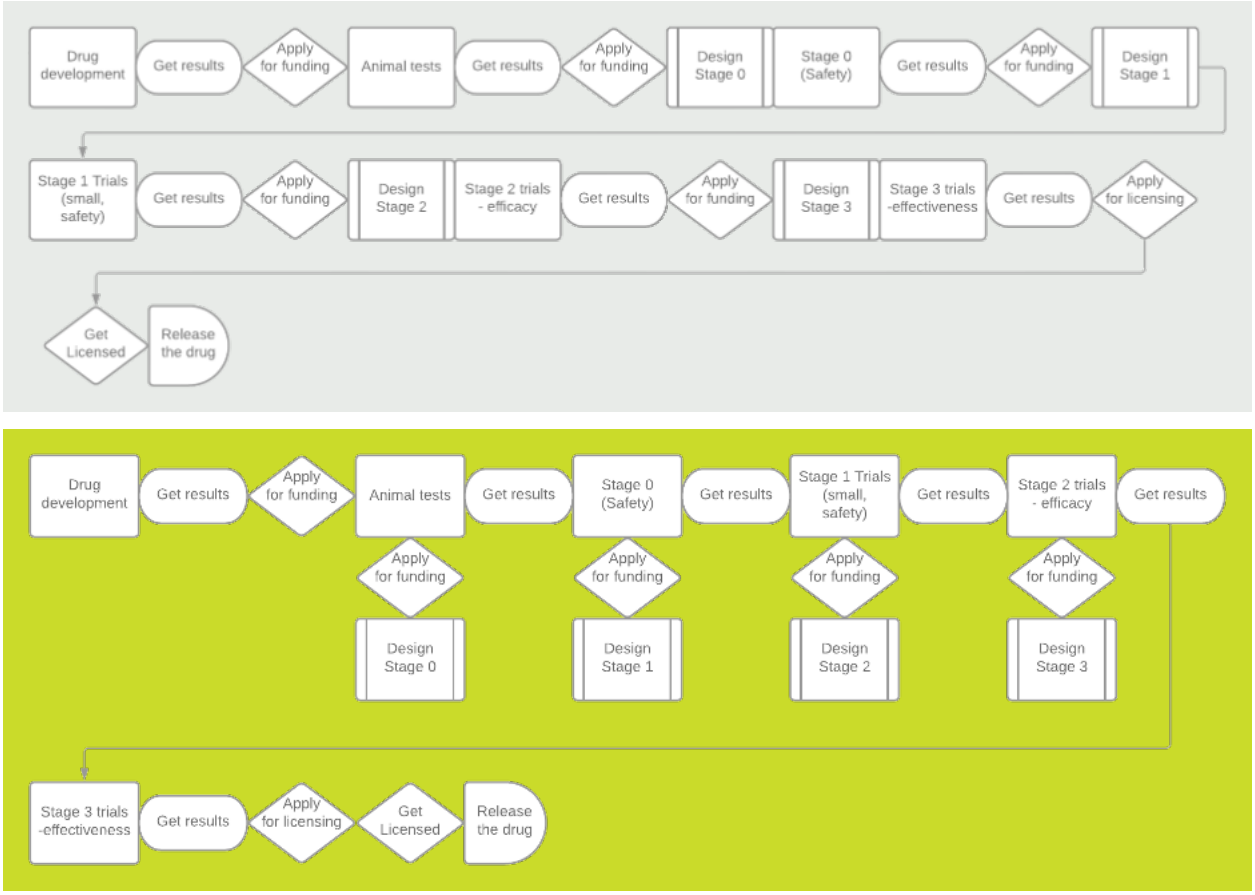
As well as proceeding "at risk", there were methodological innovations used for the first time in the development of the Oxford-Astrazeneca vaccine. One of these is to make use of a method that Gilbert developed – the "rapid method". This method involves standardising and expediting as much of the research as possible. This was facilitated by the development of a "platform" that had demonstrated safety, and on which the new vaccine could be built.

Alongside this, each phase of the research began as soon as practically possible, with preparations for the next stage happening even before the previous one had ended – see the diagram below for a description of how programmes like this usually work, compared to how they worked in this instance (see Figure 1).

This approach is risky, as time and effort might end up being for nothing. Each stage of the process acts as a gateway for the next, so if your safety trial shows serious adverse events, then you can't proceed to stage one trials. This means that any time and money you've already spent preparing for the next phase is wasted and as the phases go on, that can be a lot of money.

This approach works in a pandemic – when the need to move fast is greater than the financial constraints – and in situations where the "risk" is fairly low. Gilbert, Green and their colleagues had decades of experience developing vaccines, including against coronaviruses, and had developed a "platform" technology that overcame a lot of the basic scientific work that needs to happen when starting to develop a new vaccine. Importantly, they had a strong understanding of how the virus was likely to work – even before its genome had been sequenced – and how, therefore, a vaccine would have to work to combat it.

FIGURE 1: HOW THE DRUG DEVELOPMENT PROCESS USUALLY WORKS VS HOW IT WORKED FOR THE DEVELOPMENT OF THE OXFORD-ASTRAZENECA COVID-19 VACCINE



While much of this process cannot be applied to social policy, that does not mean that none of it can.

Why should we try to learn?

As we've said, it is often difficult to justify working at the breakneck pace that we saw in medicine in 2020. The risks of doing so – moving too fast and losing a lot of work – are extreme.

Nonetheless, we should always have in our mind the fact that our work is not without urgency. The biting urgency that necessitates a national lockdown is not with us, but we should perhaps not confuse the intensity of the remedy for the virulence of the disease.

5,980 children were removed from their parents in their first year of life during the year before the pandemic hit[2] – based on a judgement that their parents were unable to keep them safe and look after them. Overall, more than 80,000 children are in the

2 Department for Education. (2020). Children looked after in England including adoptions, Reporting Year 2020. [online] Available at: https://explore-education-statistics.service.gov.uk/find-statistics/children-looked-after-in-england-including-adoptions [Accessed 31 Aug. 2021].

care of the state. Many of these will have experienced substantial trauma, the effects of which will be more lasting, and more devastating, than long Covid. Children from a poorer family (on Free School Meals) are 33 per cent[3] less likely to achieve good grades in their GCSEs than their more affluent peers, and are 58.3 per cent[4] less likely to go to university. Beyond poverty, attendance at university for young people identified in the data as having care experience is 13 per cent, compared to 43 per cent for young people without care experience.[5]

These are just a small subset of the issues in social policy that affect thousands of people each year with long term consequences and that fall in of our particular areas of interest. Those working for our sister What Works Centres could doubtless add to this list.

The point we are driving at is not that the world is an awful place and that "something must be done". The world has been gradually improving, on average, since the end of the second world war on most of the measures that we care about. Something is being done. What instead, we are saying, is that there is a social pandemic around us, all the time. If we want to help people, we need a greater sense of urgency to our work.

What can we learn?

If we're going to learn from our cousins in medicine, it's worth outlining some of the key ways in which our methods differ from theirs.

First, our trials are cheaper. Even an expensive trial in social policy contexts will rarely cost more than a million pounds. Most education research can be conducted within the bounds of £150,000, and a quarter of a million pounds for a single trial has always been enough to give us pause.

Part of this difference in cost is driven by differences in outcome measures and in monitoring. The growth of administrative data, and its availability for use by researchers, means that we don't need to conduct expensive data collection exercises and we don't need to do anything as elaborate as drawing someone's blood, or sticking a swab up their nose/down their throat/anywhere else. Our outcomes are often things that matter to society – or at least to the government – and so they are routinely collected. This slows down our trials – we have to wait for the data to exist and to gain access to it – but reduces the cost and the administrative burden on trialists.

We're generally less concerned about side effects than medicine, and possibly more concerned with actual backfires. Our interventions are typically less likely to produce

3 Hon, R., Clegg, N., Allen, R., Fernandes, S., Freedman, S. and Kinnock, S. (2017). Commission on Inequality in Education. [online] . Available at: https://www.smf.co.uk/wp-content/uploads/2017/07/Education-Commission-final-web-report.pdf.

4 Service.gov.uk. (2018). Widening participation in higher education, Academic Year 2018/19. [online] Available at: https://explore-education-statistics.service.gov.uk/find-statistics/widening-participation-in-higher-education [Accessed 6 Sep. 2021].

5 Harrison, N. (2020). Care leavers in Higher Education: new statistics but a mixed picture —. [online] Available at: http://www.education.ox.ac.uk/care-leavers-in-higher-education-new-statistics-but-a-mixed-picture/ [Accessed 6 Sep. 2021].

side effects like blood clots, but they may be *more* likely to produce side effects in the sense of making the outcome worse. We can think of interventions in education that aim to boost attainment but actually end up making it worse, or the famous example of the Scared Straight programme, which made young offenders more likely to re-offend, not less.

Next, blinding and the use of placebos are much less common in social policy. It's impossible to meaningfully conceal from someone whether they've received an intervention in a lot of cases. Take, at its crudest, a cash transfers programme, like that trialled in Canada for homeless people, in which participants were randomly assigned to receive $7,500,[6] or not to. Neither group could be left with any ambiguity about whether they received the intervention or not. Nor can the professionals working with them. There's not a placebo that could be administered that would create this ambiguity. For this reason, combined with the more common use of cluster level randomisation, and/or uneven randomisation rates,[7] it is hard even to blind statisticians from which group is which when the time comes for analysis. In social policy, we have mostly gotten around this by not even trying to blind people, but instead, divorcing access to data and analysis from the incentive for the trial to work. Independent evaluators, not affiliated with either the funder or the intervention developer, design the trials independently of the evaluated, and then analyse the data with similar distance. This isn't perfect – as others[8] have written – but it does go most of the way.

Finally, there is a difference between what we are looking for in the early phases of testing on humans[9]. The very first trials in medicine are looking for "safety" – that is, a sense that a new drug is not going to cause major side effects, or otherwise be rejected by the body. By contrast, we are interested in feasibility. Many feasibility studies boil down to two simple questions – will professionals do the thing we're asking them to do, and will participants show up for it? In this sense, we are concerned at the outset with the acceptability of a new intervention – something that tends to be more of a question *after* efficacy has been tested for medical interventions.[10]

6 UBICenter (2020). What a Canadian experiment tells us about cash transfers and homelessness. [online] UBI Center. Available at: https://www.ubicenter.org/canada-homelessness-experiment [Accessed 31 Aug. 2021].

7 Uneven randomisation means that people are not randomised in a 1:1 ratio.

8 See, B. H., Siddiqui, N., Gorard, S. (2017). The Trials of Evidence-based Education: The Promises, Opportunities and Problems of Trials in Education. United Kingdom: Taylor & Francis.

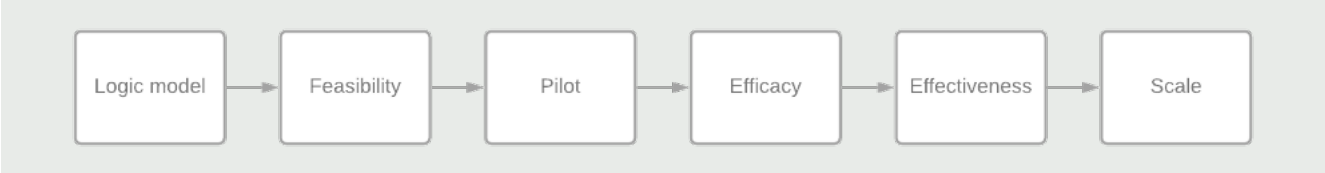
9 For social policy researchers, this phrasing seems strange – all of our testing is on humans. The knowledge that, for example, formative assessment works, or doesn't work, in mice, is of exactly no value for us.

10 Where much of 2020 was spent on vaccine development, a lot of 2021 has been spent on combating anti-vax sentiment, or vaccine hesitancy. In social policy, there is less organised resistance to interventions but it must be overcome for the trial to begin.

The "standard method" in social policy

Now, if we're going to deviate from the standard model, we should know what it is. The lack of regulation, and the heterogeneity, within the social policy sphere means that the process for developing evidence is much less well defined. But we can think of having six main phases, shown in the diagram below.

FIGURE 2: THE SIX PHASES OF THE "STANDARD" SOCIAL POLICY MODEL



Logic model

The first stage in the standard model is to create a logic model. This involves outlining what an intervention actually is, its background context and the mechanisms that, in theory, gain the desired outcomes.

Feasibility

The next step is about understanding the feasibility of the intervention working. To decide whether or not to proceed with the next steps. Does it appear that the intervention would have the outcomes it is intended to?

Pilot

The third stage is the pilot stage. This is a small-scale trial that is designed to assess the feasibility of a larger-scale trial. This stage also offers the opportunity to understand and assess how the intervention works in practice.

Efficacy

The next phase is about understanding and assessing the efficacy of the intervention. This means knowing whether it produces the desired result, or achieves the desired outcomes.

Effectiveness

While similar to efficacy, an assessment of effectiveness is about understanding the degree to which an intervention works.

Scale

The final phase in the standard method in social policy is about scale. Once we have gone through the proceeding stages, we will know whether an intervention could be scaled up successfully or not.

The rapid method

Where, then, can we meaningfully move faster and save time? There is much to learn from the "rapid method" that the Oxford team developed and deployed. We could recognise, for example, that many of the stages of our traditional research pipeline can be carried out at the same time. A feasibility study can be conducted as the first stage of a randomised controlled trial – if an intervention isn't feasible, it will fail fast, and the trial can stop. The pilot stage – where we evaluate whether mechanisms that should, in theory, change as a consequence of our intervention are changing – similarly can be tested in real time, while a trial is going on.

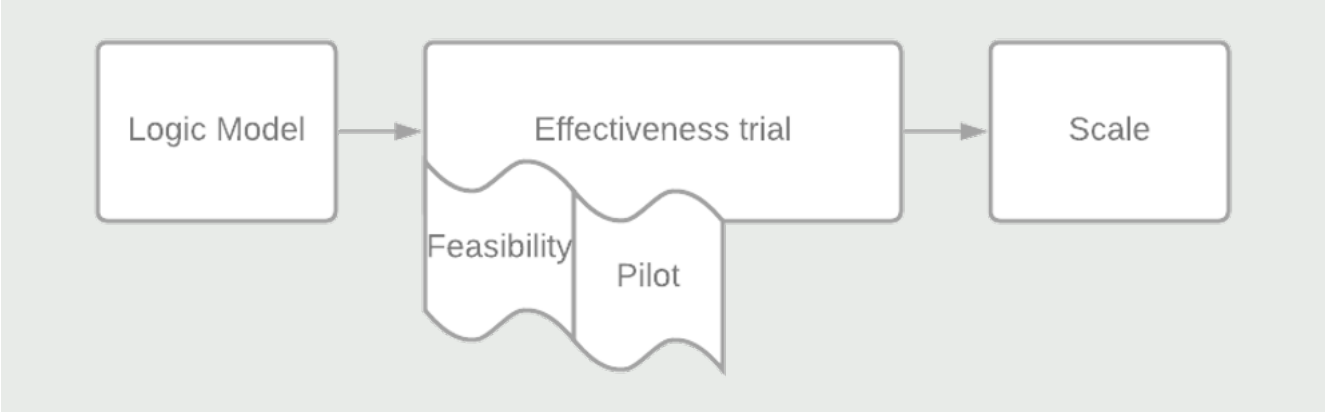
If the mechanisms don't change, and you're confident that the mechanisms are the way that the main outcomes will change (we often aren't), we can discontinue a trial before reaching the main outcome stage. Given that in social policy we don't have anything so neat as antibody production to tell us that we're heading in the right direction, we may opt to continue anyway.

There is also a falsehood in the distinction between efficacy and effectiveness trials. What are ideal circumstances in social policy? Is an intervention most likely to succeed in an outstanding local authority with strong leadership, a stable workforce, and excellent existing services? Or is it most likely to succeed somewhere that's at the beginning of an improvement journey, having recently been rated as inadequate by Ofsted? In the former case, maybe not because the services are already maximally good, so achieving improvement is likely to be hard. In the latter case, where there is a lot of benefit to be gained, it may be difficult to get traction for *any* intervention. If you think I'm drawing a straw man here, I've heard both arguments from directors of children's services, in the last month, about the same intervention.

If we can't tell what ideal conditions are then the role of the efficacy trial as a gateway to an effectiveness one makes no sense. We potentially abandon impactful interventions too soon, or proceed with ones that have no hope of working at scale. Instead, our rapid method should proceed straight to an effectiveness trial.

If we take all of these things together, what are we left with? A model that looks like the diagram below, combining multiple phases into one, with gateways after each. By doing things this way, we could cut certainly months, but perhaps years, from the process of evidence generation.

FIGURE 3: THE RAPID MODEL



Conclusion

This short paper has outlined a case for a rapid method for social policy trials. Under this new model, real-world delivery, with randomisation and collection of baseline and endline data, would begin as rapidly as possible after an intervention had been fairly well defined through a logic model process.

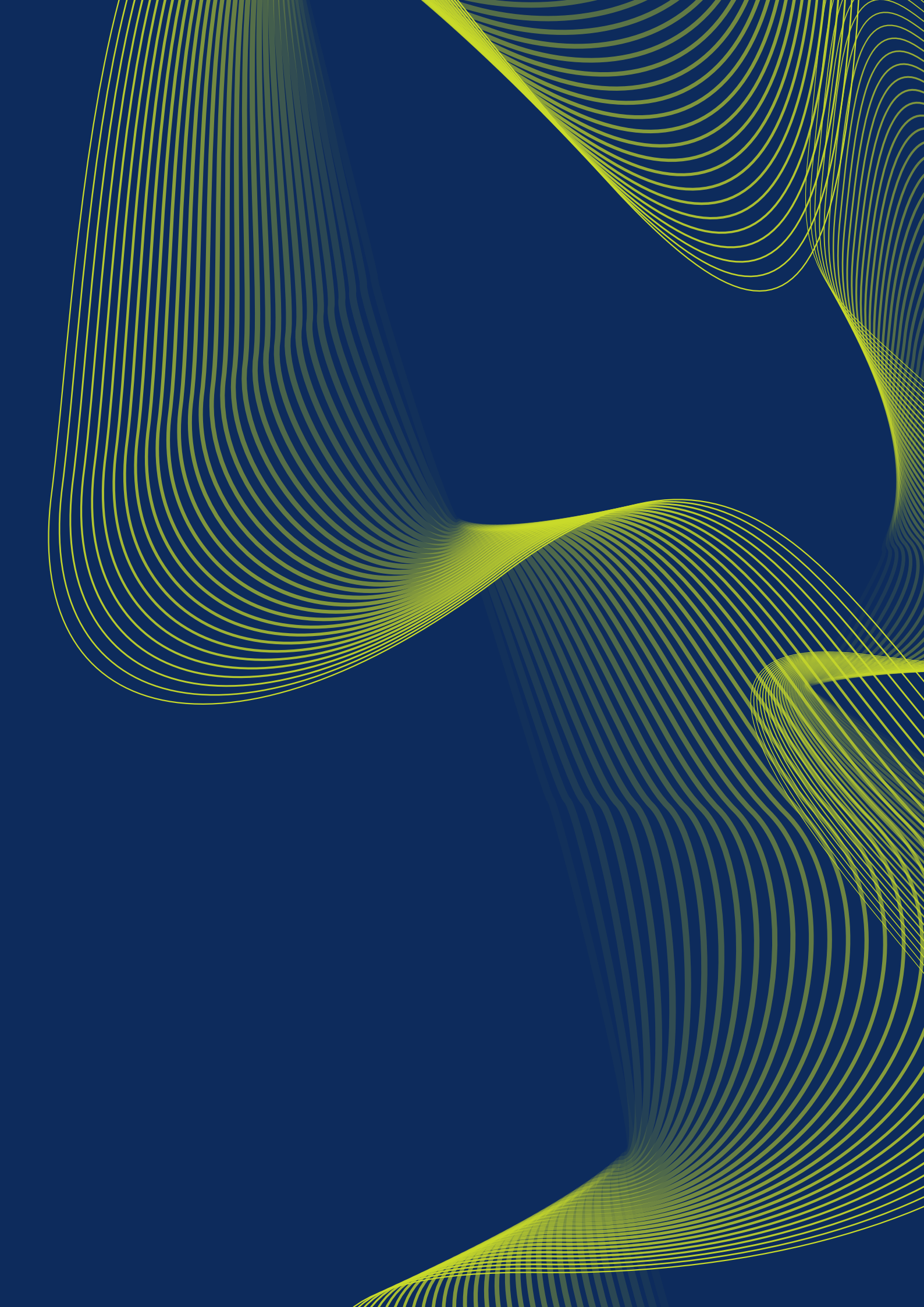
After the feasibility stage of the trial – that is, after a set amount of time had been passed and a certain number of participants had been recruited, consented, randomised and had the chance to participate (or, indeed, not to) – the trial could then be continued or not.

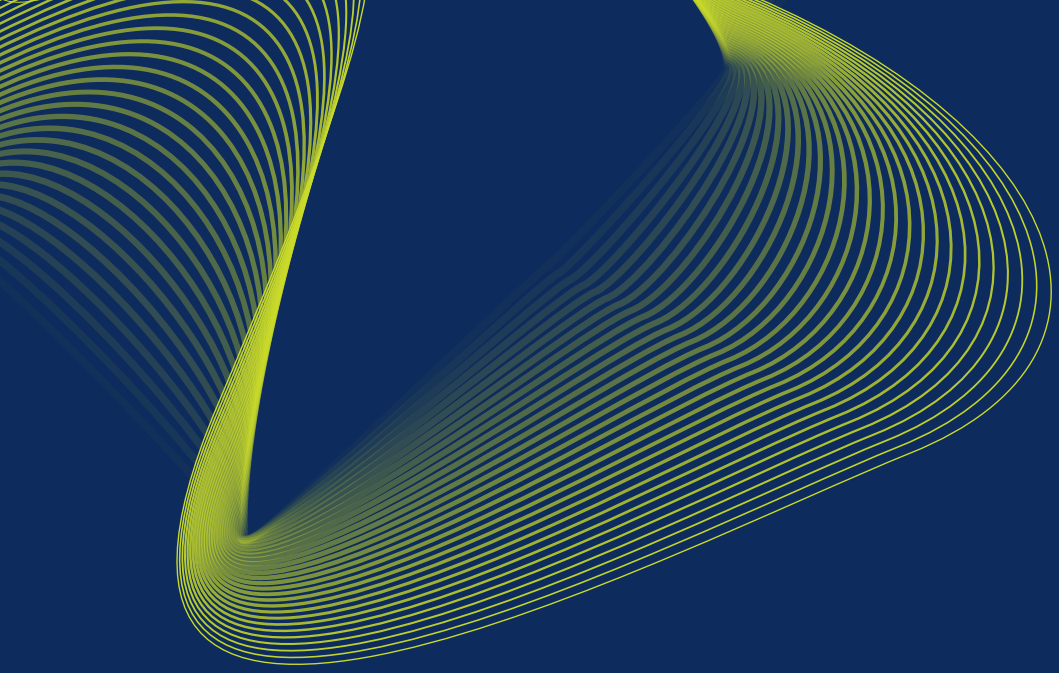
Implementation and process evaluation – looking to see if important mechanisms were changing, whether or not professionals and participants found the intervention valuable, and so on – would begin from the outset and plausibly draw conclusions within six months of a trial beginning. This being done, a decision could again be reached as to whether or not to continue.

Importantly, randomisation and consent for data capture would begin at the very beginning of this process, meaning that all participants throughout these phases could be included in the trial's analysis, and that the time taken on standalone feasibility and pilot studies can be saved. This could potentially save years.

There are practical considerations to work through though. How would evaluators, many of whom are commercial concerns, handle the uncertainty of these trial gateways? How would ethics committees view randomisation of early stage participants in a trial that may not reach its endline point? It is important that these questions, and others, are answered before adoption of the new method.

However, it is fairly clear that in the status quo approach – where interventions can take half a decade to reach trials and many promising interventions are rolled out after feasibility or pilot studies due to political expedience – evaluators and other advocates for What Works are not moving as fast as we can to tackle the very real problems we are committed to fighting. Even if this new method does not work, we must find some way of working faster.





The Policy Institute

The Policy Institute at King's College London works to solve society's challenges with evidence and expertise.

We combine the rigour of academia with the agility of a consultancy and the connectedness of a think tank.

Our research draws on many disciplines and methods, making use of the skills, expertise and resources of not only the institute, but the university and its wider network too.

Connect with us

@policyatkings kcl.ac.uk/policy-institute