The spread of severe acute respiratory syndrome coronavirus has thrown the world into crisis. Public health responses have been prolific, but the success of these interventions is uneven. As ever, we need to know what works for whom in what circumstances and in what respects. Realist approaches have been devised to answer this question and this working paper series is designed to promote this cause.

*The series offers immediate readership and has no formal refereeing process. They are working papers in the sense that they are intended to raise hypotheses leading to more productive evidence. They may be developed (or indeed withdrawn), they may go on to be published in journals and books. Above all, they are attempts to provoke dialogue in the realist community and beyond.*
THE RELEVANCE OF REALISM IN THE PANDEMIC


Author: Ray Pawson
Date: DEC 2020

A Complex Coronavirus Brainteaser:

Has the interlocking, mutually dependent and sometimes competing stockpile of adaptive, self-transforming, interventions against the virus, each one with complex and sometimes contested guidance on its remit, as implemented and switched on and off by a changing array of central, local, private agencies, as shaped by competing political demands and factions, as digested by a diverse population containing people who variously support, comply, prevaricate, resist, grow weary, change their minds, and seek exceptions, succeeded in controlling the virus?

All contributions in this series may be considered ‘open access’. Please feel free to circulate and to quote from this paper. If you do make a reference to it, and as a courtesy to the authors, please let them know.
The Denial of Complexity in Developing the UK Covid-19 Response.

Ray Pawson

Abstract

There is a popular conception that experts are experts because they have access to privileged knowledge. The appliance of science provides them with authoritative evidence. But scientific consensus is notoriously hard to achieve. When it comes to remorselessly complex problems like designing policy to combat the spread of the coronavirus, experts rarely speak with one voice. A question thus arises about the hierarchy of expertise consulted on crisis policy – which voices get a hearing, and which voices go unheard? The paper examines this issue by charting the differential access of a variety of scientific disciplines to the UK Scientific Advisory Group for Emergencies (SAGE) during the first year of the Covid-19 crisis. It turns out that truths are not self-evident and not all disciplines are created equal. The epidemiological, clinical and statistical sciences dominate this executive. A particular casualty of the pecking order is the lack of attention paid to the notion of system complexity; a perspective widely employed in the evaluation of major social programmes and policies. The paper draws attention to the evidence overlooked in mainstream epidemiological inquiry and concludes with a manifesto for an alternative, complexity-sensitive research programme.

Key words: coronavirus, complex adaptive self-transformative systems, Scientific Advisory Group for Emergencies, policy evaluation

First draft for commentary and criticism.
Revised version due to appear in:

Introduction

It is often proclaimed that the policies and programmes selected to mitigate the Covid-19 epidemic will be ‘led by the science’. This laudable objective raises intriguing questions. There are many branches of science and attempts to classify them run literally into hundreds of types and sub-types (Pierce, 1902). So, which of them should be represented at the evidential top table and how is the selection made? More specifically, and here we come to the particular concern of this essay – to what extent does social inquiry figure in the higher echelons of research expertise and what is the nature of its contribution?

The paper is structured as follows. In the UK, the top table in question is the Scientific Advisory Group for Emergencies (SAGE) and the first section examines its membership. A notable absentee is identified, namely social scientists involved in the study of complex systems. The second section examines the nature of policy response to epidemic, revealing that it is a complex system par excellence. Accordingly, much of the statistical and epidemiological evidence considered by SAGE fails to grasp the adaptive, self-transforming nature of the societal response to virus policy. The third section provides examples of complexity-based research on mitigating virus transmission, demonstrating its potential to provide evidence omitted by mainstream epidemiological methods. There is a short conclusion.

Who’s who in UK coronavirus research?

Some acronyms are ingenious, none more so than the notion that government action should be based on SAGE advice. The Scientific Advisory Group for Emergencies is the standing committee that provides crisis management advice to ministers on all major national emergencies. It does not operate under government instruction, although government officials are represented at its meetings. Participants are selected by the British Government Chief Scientific Adviser and the Chief Medical Officer for England. For the coronavirus crisis, the standing committee was supported by several sub-groups as listed in Table 1.

Table 1: The SAGE committee structure in response to coronavirus outbreak (January 2020)

| Scientific Advisory Group for Emergencies (SAGE) |
| Scientific Pandemic Influenza Group on Behaviours (SPI-B) |
| Scientific Pandemic Influenza Group on Modelling (SPI-M) |
| PHE Serology Working Group |
| COVID-19 Clinical Information Network (CO-CIN) |
| Environmental Working Group |
| Children’s Task and Finish Working Group |
| Hospital Onset COVID-19 Working Group |
| Ethnicity Subgroup |

Notes. SAGE also uses ad hoc evidence supplied by other bodies, most significantly the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG). SPI is stands for ‘Scientific Pandemic Influenza Group’ – an acronym retained from previous crises.

This table provides an immediate empirical answer to our question on the selection of scientific expertise; it provides, so to speak, a formally sanctioned ‘hierarchy of evidence’. What voices are represented in this structure and what are their methodological orientations? Group membership is
listed on the government website (unless the participant prefers to remain anonymous), so it is reasonably easy to profile the respective disciplinary backgrounds. As might be expected the main committee is a fairly broad church, with some crucial exceptions. The clinical sciences are well represented with physicians, virologists, immunologists, microbiologists and epidemiologists. There is a great deal of political cover with representatives from various ministries and the devolved powers. The formal sciences constitute another major bloc, via statisticians, big data specialists and mathematical biologists. Note, that the latter group, the epidemic modellers, also have their own substantial sub-committee (SPI-M), giving them two bites at the cherry. Thereafter, there is a scattering of individuals from behavioural science, the law and business.

We turn next to SPI-B, given our interest in social science’s representation in the advisory elite. Who sits around this table? There is a curious imbalance. The majority are psychologists, behavioural scientists, health informaticians, health communicators, etc. No less than three members hail from the Behavioural Insights Team (a.k.a. the ‘nudge unit’). There are very few sociologists, political scientists, management scientists, or economists. Most curious, given that the basic response to the epidemic took the form of an unprecedented exercise in social control, is the lack of policy analysts, programme evaluators, complexity theorists and systems analysts. It is unclear how this committee was assembled. But, to put it politely, it appears more of a snowball sample than a representative sample or even a purposive sample of the appropriate expertise.

Nothing decisive, of course, can be gleaned from the disciplinary make-up of these committees. Recall that their role is to offer advice. It remains Government’s job to formulate policy. We can, however, move closer to precise scientific imprimatur behind the chosen advice by looking more closely at the remit and recommendations of key groups, namely I. SAGE, II. SPI-M and III. SPI-B

I. SAGE. A vast amount of evidence emanates from the main SAGE committee, in the form of meeting minutes, background papers and formal publications. The 2020 outputs can be summarised as follows. The primary function is to chart minutely the ‘vital signs’ of the influx of the COVID-19 virus. Real-time data is collected on prevalence, regional spread, the R number, incubation periods, hospitalisation rates, ICU usage, death rates and so on. This work draws on the considerable resources of bodies such as the Office for National Statistics (ONS). Other research is consulted on the adequacy of resources such as masks, beds, ventilators, protective equipment, back-up facilities (Nightingale hospitals), testing-systems and laboratory capacity (the Lighthouse Labs), training requirements, and so on. Advances in the treatment (e.g., Dexamethasone) and the management of patients with the virus are also scrutinised. A waiting brief and watchful eye is also kept on the longer-term prospects for salvation in the form of an effective vaccine.

The committee then turns from evidence to guidance. On the basis of this medley evidence, the policy options are considered, advice is prepared on the likely impact of a range of potential interventions, and when to turn them on and off. These measures came thick and fast as the virus deepened: hand hygiene advice, surface cleaning, social distancing, mask wearing, the closure of shops, stadiums and schools, isolation and lockdowns, working from home, testing and tracing, travel bans, the rule of six, family bubbles, tiered regional responses, and so on. This is the stuff of science, making causal inference, assessing impacts, deciding what works. SAGE’s modus operandi is

---

1 Other authors have made a much closer study of the composition of the SAGE committees, searching for key influencers and conflicts of interest (Harcombe, 2020; Thacker, 2020)
to ‘track and dose’ – to measure the extent and spread of the epidemic at any point and then to recommend a particular package of interventions as the appropriate level of response. In the next main section, we consider in more detail the adequacy of these inferences. For now, we simply identify the moment when evidence turns to policy. SAGE advises and Government ponders.

II. SPI-M. The role of the modellers is also crucial. The voice is undeniably loud. This group deal with the same body of evidence just described but with the twist of propelling the impact assessment into the future. How many people, with what institutional support, will self-isolate, work from home, take up the furlough scheme, obey travel restrictions, abandon public transport, trade less, keep to the 2-metre rule, commit to home schooling, modify family ties, establish social bubbles, and so on. No one knows the answers to these questions in advance of policy implementation. But estimating the extent and effectiveness of such social change is precisely the task entrusted to the model builders. This is no place to describe the methodology. Suffice to say that the basic process (Keeling and Danon, 2009) is to make estimates of the levels of adherence to these reforms, to gather these many assumptions together statistically, in order to produce graphs quantifying the difference between ‘doing nothing’ (cases without intervention) and ‘doing something’ (cases with the interventions that might put in place). Under crisis conditions, the wisdom governments seek is about the future, they want assurance that upcoming policy will work, and the extrapolations of SPI-M were entrusted to pronounce on just that. In the next section we examine whether that trust was justified.

III. SPI-B. We now turn to the behavioural scientists and begin with their remit: ‘SPI-B provides advice aimed at anticipating and helping people adhere to interventions that are recommended by medical or epidemiological experts’. There are several points to note here. The first is the jaw-dropping deference to the polymaths who reside elsewhere. Expert medics and epidemiologists are entrusted with researching and recommending the interventions; behavioural sciences are there to promote policy and nudge the public into compliance. Messaging is SPI-B’s mission. As each extension of provision is mooted, the sub-committee, ‘reviews the interventions in terms of public attitudes and support; likely adherence; and any barriers, facilitators or communication issues that should be considered’ (GOV.uk, 2020). There is no space to assess these recommendations here, other than to remark that they have ‘common sense’ appeal - make sure that guidance is clear, provide a rationale rather than assert rules, consult on messaging details with key practitioners, and so on. Each recommendation has a theoretical pedigree and can be traced back to key principles of behavioural motivation: judgement heuristics, cognitive dissonance theory, temporal discounting and so on (West et al, 2020). But note again the deafening silence in expectations. Social science has no role in selecting and designing interventions; the task is to help with their comprehension. In the final section I will argue for a revised and fuller job description.

This completes an initial review of the divisions of expertise and labour in the SAGE committee structure. Other groups provide much more specialised contributions. The Serology committee reports on the diagnostic examination of blood serum with regard to the immunity responses, the Ethnicity committee advises on risks and impacts for minority groups, and so on. The paper now turns to the key question – what expertise is actually required to evaluate the totality of the policy response to the virus, has the correct scientific division of labour been struck, and should social science be limited to the duties of the public relations advisor.

The policy response as a complex, adaptive, self-transforming system

Perhaps the most elementary rule of scientific inquiry is to fit research method to subject matter. Investigation begins, often invisibly, with ontological assumptions on the nature of the object of
investigation, its basic properties and the relations between them. So, what is the nature of the first wave of interventions to combat the coronavirus pandemic? Whilst the world holds on to the promise of an effective vaccine and better clinical treatments, the entire first year of pandemic policy has taken the form of what SAGE minutes usually refer to as ‘non-pharmaceutical interventions’ (NPIs). This is a curious and completely uninformative designation for it tells us what the response is not. If we turn to the published outputs of SPI-B, the preferred designation is ‘behavioural interventions’. Again, it is a spurious label. To be sure, any reduction in transmission depends on people modifying their everyday conduct but behaviour change is hardly the sum total of government policy.

What the coronavirus response actually consists of is a massive array of public policies and social interventions that reach into every institution and every sphere of public and private life. It is an unprecedented attempt to manage social change on a grand scale. It is, in short, a complex, adaptive, self-transforming system. There has been a significant ‘turn’ towards complexity and systems thinking across the social science in recent years (Williams, 2021) and this orientation features increasingly in the evaluation of national programmes and policies (Gates 2016). The properties of complex systems have been dissected in detail (adaptation, emergence, unanticipated consequences, feedback loops, blockage points and structures, non-linearity, tipping points, path dependency, openness and self-transformation). This is no place to explicate these abstract processes, but it is simple matter to demonstrate that they are deeply implicated in the interventions designed to combat the virus. Herewith, a summary of those manoeuvres (Box 1).

Box 1: The COVID-19 policy response ...

1. ... Consists of an ever-mutating bundle of interacting programmes (hand hygiene, protective equipment, closure of shops, stadiums and schools, social distancing, rule of six, working from home, testing and tracing, lockdown 1 and 2, etc.). Each intervention conditions the others, often in emergent, unanticipated ways. Discharging the elderly without testing reduces hospital infection but increases care home transmission. Isolation measures decrease A&E loads, cancer referrals, pollution levels but increase mental health problems, domestic abuse, educational disadvantage. ‘Eating out to help out’ revives economic activity but reintroduces virus transmission. Vaccine announcements induce complacency, and so on.

2. ... Involves long implementation chains, which adapt the interventions on their way to the public. Central government or ‘top down’ interventions often come with pages and pages of guidance, which are reinvented continually by intermediaries over time, generating diverse consequences. Hospitals and care homes extemporise under demand pressures, PPE shortages and staff absences. Policing policy on unofficial gatherings varies by constabulary. Schools differ in maintaining provision for the at-risk and the children of key workers. Parents disagree about safety levels on reopening. Family’s resolve to isolate and distance weakens over time, and so on.

3. ... Requires as much if not more attention to ‘exit’ as it does to ‘entry’. Unlocking is significantly more difficult to phase, manage and implement than lockdown. Closing schools, shops, stadiums and so on is much simpler than reopening them with capacity limitations, one-way systems, sanitising points, screening and booking systems. The introduction, relaxation and reimposition of controls (lockdown 1 and 2, regional and tier systems, firebreaks, Christmas break etc.) enrage business and confuse the public.

4. ... Is deeply contextual, with the same measures generating different outcomes in different communities and countries. Both the transmission potential and the capacity to respond vary significantly from location to location. Disease prevalence varies significantly by subgroup. The $R$ number varies sharply from neighbourhood to neighbourhood. Compliance with guidance varies with national and local culture. Public health discipline changes. Very young children, dementia sufferers...
and the drunk and disorderly have little capacity to obey distancing rules. Guidance is continually tested by 'free riders' and so on.

5. ... Is continually buffeted by political dogfights, with frequent changes in strategy and in action plans. The timing of the introduction and withdrawal of specific interventions is influenced almost daily by media and social media exhortations. The UK response has twisted and turned under pressure from powerful libertarian and authoritarian interest groups (and celebrity footballers!). Most significantly, the content and substance of the response has been shaped and reshaped under negotiation with professional bodies, with local authority leaders, with the different national jurisdictions and so on.

6. ... Consists of a complex, adaptive, self-transformative system, thrust into a complex, adaptive, self-transformative system. All policies, including those directed at epidemics, operate in a wider cycle of reforms. Coronavirus interventions shape and are shaped by other contemporaneous social movements and political agendas. Harsher restrictions impact mainly on deprived areas in the north of England and thus creates tensions with the so-called 'levelling-up' agenda. The virus outbreak coincides with the Black Lives Matter movement – and high incidence rates in certain communities lead to charges of 'institutionalised Covid'. Brexit and ecological concerns lead to further tensions.

7. ... Uses social interventions to break biological chains of transmission that are often unknown and largely invisible. Estimates for the percentage of people testing positive for SARS-CoV-2 who may be asymptomatic vary wildly (between 5% and 80%). Interventions involving symptom-based screening will inevitably miss cases. The virus, moreover, is a syndrome consisting of several mutating strains which vary in their origin, prevalence and stability. Over time these may generate novel and unanticipated threats.

Complexity abounds. In each category, I have provided only three or four illustrations of the particular system dynamic from the scores of parallel examples that could have been chosen. What are the methodological implications? What limitations are placed on 'the science'? This brings us to the core contention of the paper, namely that the linear, additive assumptions built into SAGE’s tracking and modelling methodologies leave it poorly equipped to evaluate the distinctly non-linear and emergent developments in virus topology. The central concern is this – causal attribution is the classic aim of applied science and drawing causal inferences within complex systems is extremely taxing.

As with all policy evaluation, coronavirus research tackles the ‘what works?’ question. Put simply:

*Have the assembled interventions succeeded in controlling the virus?*

If the above analysis is correct, the causal question is transformed thus:

*Has the interlocking, mutually dependent and sometimes competing stockpile of adaptive, self-transforming, interventions, each one with complex and sometimes contested guidance on its remit, as implemented and switched on and off by a changing array of central, local, private agencies, as shaped by competing political demands and factions, as digested by a diverse population containing people who variously support, comply, prevaricate, resist, grow weary, change their minds, and seek exceptions, succeeded in controlling the virus?*

This ugly circumlocution provides a pellucid summary of the real task faced by the scientific advisors. Traditionally, causal analysis in non-experimental (e.g., public health) studies seeks statistical association by tracking the application of an independent variable against a shift in a dependent variable. This approach, with the inclusion of a few mediators and moderators, can work reasonably well when intervention and outcome are simple and singular (e.g., clean water supplies and diarrhoeal deaths). But when the independent ‘variable’ is actually a continually mutating
conglomeration of interventions and when the dependent ‘variable’ is a whole dashboard of fallible indicators, identifying the active causal pathways becomes a methodological minefield.

So how does SAGE go about mine clearance? The vital signs of the epidemic are monitored in great detail on a daily basis (prevalence, deaths, hospitalisations, ICU capacity, R number, etc). This aggregate picture provides an immediate but approximate measure of the level of control over the virus at any point in time. When control is perceived as lacking (‘the virus is spreading exponentially’) a tougher medley of measures is recommended (lockdown 1 and 2, tier 3 controls, etc.) and when control appears to gain ground it is suggested that selective measures may be loosened (COVID-secure opening of schools, pubs, shops, etc). The crucial point here is that the judgements and adjustments exercised here are just that. We might dress them up as ‘expert opinion’ but there is no exact formula, no statistical algorithm available to say that a particular configuration of transmission data demands a particular configuration of response – a conclusion that is hardened when we recall that politicians rather than scientists make the crucial calls.

The ultimate question is how do we know that a chosen package counters the varying prevalence of the virus at any point in time? This is where the public expects science to pronounce in but, because of the complexity of inputs and outputs, it is no ‘exact science’. What is the basis of such decisions? Most of the available interventions rely on a simple mechanism – reducing social contact reduces virus transmission. Some elementary causal reasoning is then applied – as the virus lifts apply more movement restrictions and as it falls take some of them away. This is something akin to steering a huge tanker in rough seas. We know when it is wise to tack to port and when to starboard but not exactly by how much. Worst of all the steerage takes place in slow motion and takes effect long after the wheel is turned.

In short, it is perfectly possible to chart the ups and downs in the prevalence of the virus and to know when action is due, but it is not possible to presage the specific and delayed effects of a particular package of restrictions. What we cannot do, because the interventions are mutating, quarrelsome conglomerations, is to perform ‘contribution analysis’. We cannot pinpoint the crucial causal agents and their specific effects. We cannot say that ‘handwashing to the tune of happy birthday’ contributed to a% of the reduction in transmission, that ‘school closure’ accounted for a further b%, that the ‘track and trace’ generated another c% improvement, nor that ‘eating out to help out’ caused a rebound of d%, or that the ‘Christmas truce’ will add e%. And so on. We cannot say any of this because, as complexity theory teaches us, each of these contribution transforms continually in the presence of all the other system dynamics.

In this respect it is also important to challenge the casual use of language which says that ‘lockdowns’ work. Lockdowns are agglomerations. Lockdown 2 was not the same as lockdown 1, nor is it the same as the lockdowns that have been tried across the world. To be sure, total lockdown (Wuhan style) might well work but only when applied unsparingly in a highly regimented and compliant population. The task in the UK remains that of tailoring and balancing lockdown’s many components in order to anticipate the changing response from its many publics. Trial-and-error is inevitable.

What should be even more obvious is that it is impossible to propel estimates of the impact of specific interventions into the future. As we have seen much of the early scientific advice came from the phalanx of mathematical modellers within the SAGE committee structure. The modeller’s task is to make graphical projections on the difference between ‘doing nothing’ and ‘doing something’. In order to produce these estimates, assumptions have to be made about none other than the ‘contributions’ of the various micro-components in the interventions. Such ‘parameterization’,...
because it is conditioned by the statistical capabilities of the models, always tends to the fixed, mechanical and linear – x% of population will comply with measure A; y% of hospital patients will require ICU treatment, z% of schools will close and so on. In a previous paper I have scrutinised the assumptions built into models that were particularly influential in SAGE advice and found them severely flawed (Pawson, 2021). The challenge to statistical soothsayers is quite simple: contemplate the continuously adapting and self-transformative complexity outlined in Box 1 – and model that!

I conclude that the mainstream UK response, the national ‘track and change dosage’ model works only up to a point. It provides solid, broad-brush advice on when to ratchet up and when the ratchet down provision. But it can never grasp the level of granularity required to predict or follow every micro-circuit of transmission occurring in a complex system (Manzo, 2020). It will forever run into counterclaims which assert that a particular national or regional intervention is unwarranted and unfair because its blanket impositions fail to recognise significant local or ‘post code’ differences. The attempt to ‘strike a balance’ will never be tolerated by those committed to different economic, educational, health and welfare balances. Failures will always occur, though they can only be recognised retrospectively (the sluggish onset of UK restrictions, the erroneous early mathematical models discounting the threat from large sporting fixtures, failure to anticipate massive care home mortality, the premature easing of lockdown 1, and many more). In summary, national policy has and will continue to meet with mixed fortunes.

Note well – this is not meant to be a particularly censorious criticism. What I am describing here is a conclusion reached in many decades of inquiry on government decision making under conditions of duress. This research always harks back to a paper by Lindblom (1959), with the unprepossessing title: ‘The science of muddling through’. He suggests that it is impossible to master the complete circuitry of truly complex social problems. More modesty and increased scepticism is required from the research community on that score (Merton, 1942). Not everything can be anticipated and, like it or not, constant adaptation and revision (and thus muddling) is inevitable.

The positive suggestion, however, is that policy makers and researchers should shift more attention down to the level of the component interventions and the incremental steps that might be put in place to make them work better. Under such a purview, programme implementation becomes a priority and centralised expertise gives way to devolved decision making. Rather than relying entirely on expert opinion on the precise concoction of interventions that will overmaster the virus, the neglected option is to target more research resources on managing, improving and auditing those interventions, which have found their way onto the table.

Policy research in complex, open systems

There is a paradox at the heart of complexity theory, expressed succinctly in the two epigrams at the start of this paper. Firstly, unless we admit complexity into our thinking on the virus, we fall prey to the tendency to trust preordained solutions, or worse the Trumpian tyranny of denying that there is a significant threat. On the other hand, complexity can be seen as the enemy of execution. To acknowledge the sheer impossibility of charting every consequence of the totality of measures in a self-transforming system is to invite policy paralysis.

Both of these dramas can be avoided as soon as we recognise that we have always been citizens of a permanently complex world and that we often managed to progress tolerably well on the basis of partial, fallible evidence. This maxim applies to science itself. Any student of Popper recognises that it is an evolutionary exercise, in which we tackle problems with tentative theories. These theories
always meet with success and with error. Science progresses and progresses well by revising theories with the gradual but never-ending task of error elimination (Popper, 1979).

Rather than seeking to perfect the whole suite of responses to the virus, rather than muddling and meddling the entire repertoire, another approach might be to concentrate more on some orthodox methods of policy research namely, scoping, planning, implementing, improving, and evaluating interventions. This might sound blindingly obvious to the policy research community but remember that community is not well represented in the SAGE hierarchy. SAGE utters advice on which interventions to pursue and how to publicise them – but not on how to implement them. That task is entrusted to others.

In the past decades, a huge repertoire of methods has been developed under the badge of policy and programme evaluation. Here I want to concentrate on what might be termed formative or developmental approaches. Evaluation, quite properly, is often considered an ex-post strategy, determining after the event whether a policy or programme has worked. But that is not SAGE’s remit, nor is hindsight a wonderful thing in the middle of an ongoing epidemic. What strategies might provide foresight or at least real-time insight? There is huge toolbox of methods of ex-ante evaluation and I have space here to offer three brief examples in order to illustrate their potential value in formulating virus control policy.

1. **Concept mapping.** The first is often termed concept mapping (a.k.a. ‘logic modelling’, ‘theories-of-change’, or more simply ‘logistics’). In interventions with lengthy implementation chains (see Box 1, point 2), the requisite steps in the pathway are first mapped in great detail. Consideration is then given to the blockages, challenges, unintended consequences that might befall each stage. These concerns are gathered from previous experience and earlier inquiries into similar programmes. ‘Implementation science’ then grows as a repository of knowledge on the potential flows and blockages within particular families of interventions. On this basis a ‘delivery model’ is the standard requirement of major public policies – an implementation chain is devised, it is populated with an infrastructure of providers and resources, with the associated risks charted for each link in that chain.

   Let us consider the merits of this strategy in relation to a major component of the virus response, known in the UK as the National Health Service ‘Test & Trace’ programme (NHST&T). The basic logic, variations of which are used worldwide, is to identify those individuals with an infectious disease and to trace their contacts and then to encourage these contacts to self-isolate. The brave ambition is to identify and then break active micro-circuits of transmission. NHST&T was launched in May 2020, with a budget of £22 billion.

   Notoriously, ‘the government did not document the basis for the delivery model it chose for the national test and trace programme in a business case until September 2020’ (National Audit Office 2020). This delay led to much political suspicion and consternation (e.g., ‘Outsourced and undermined: the COVID-19 windfall for private providers’). But our concern here is for formative and developmental evaluation and the significant but bizarre consequence that much of the public scrutiny of NHST&T implementation was carried out by external bodies such as the National Audit Office, the Health Foundation, and indeed by investigative journalists.

   Their usage of concept mapping approach turns out to be most instructive as illustrated in figure 1. Note that is particular map was provide by BBC journalists extracting information from contractual records. It demonstrates at a glance the complexity of the intervention – initiated by the civil service, bolted together rapidly, then controlled by a portmanteau of private firms.
What might an implementation science perspective have to say about such a structure? There are some general rules of thumb: build on existing capacity, use experienced providers, avoid overcomplicated chains of command, provide end-to-end surveillance. Whilst it is perfectly true that no existing infrastructure was up and ready to perform such a mammoth exercise, the chosen delivery model did mean that much local expertise available in the NHS, Public Health England, the universities and local councils was initially bypassed. A catalogue of complications and shortcomings followed – a small sample of which are discussed shortly. The organisation upshot, predictable enough however, was that by the final months of 2020 a revised programme of local and mass testing was returned where it belonged, namely to health professionals.

Each part of Figure 1 can be mapped in much greater detail and if we turn to contact tracing, the logistics involve another convoluted chain (not illustrated) involving rules on: the number of contact tracers required, their recruitment and training, number of attempts and time allowances to reach people with positive tests and their contacts, scripts for call handlers, different tiers and approaches for hard-to-reach cases, identifying support needs of different contacts, and so on. The effectiveness of the scheme depends on the veracity of each step in the logic model, on which the NAO (2020) reports in detail. I quote here the evidence on utilisation rates of contact tracers (the utilisation rate is the amount of utilised time in hours divided by the number of paid hours).

‘On 17 June, following the launch of NHST&T, the utilisation rate of contracted staff was just 4% in Tier 2 and 1% in Tier 3. The Department’s analysis indicated that part of the reason for the under-use was the lower-than-expected numbers of cases and contacts being handled by the tracing service. It also recognised that its initial planning assumptions had been incorrect. For example, the Department had assumed that each case transferred to the tracing system would provide 10 to 30 contacts (the actual number was 2.4) and that 10% of cases would return details via the online route (the actual proportion was 20%).’
The NAO report concludes drily that, ‘levels of under-utilisation mean that substantial public resources have been spent on people who were barely used in actual tracing activity’. The reaction of the press and public was somewhat more hysterical. A proper ex-ante evaluation would have advised on building to scale, rather than scaling to budget.

Digging deeper one notes that the aforementioned contact tracers for the main service were hired by outsourcing companies. Health service experience was not a requirement as they worked with a closely scripted interview2. What might a thorough ex-ante assessment have concluded? Two vital pieces of foreknowledge might have been applied. A recent Cancer Research UK (Rosenberg et al, 2019) study notes that verbal advice on smoking cessation is much more likely to be heeded if delivered: i) at a teachable moment, by ii) a primary care practitioner, who is iii) able to refer patients directly to stop smoking services. Another, more immediate parallel is the UK NHS 111 service which uses call handlers to provide non-emergency support advice. On instigation it was on the end of similar hostilities about the use of non-qualified staff, and it took many years and a significant research effort to create the infrastructure to balance patient care, medical expertise and workforce reconfiguration (Turnbull et al, 2014). In both cases, structures rather than scripts are the requirement.

As a final example, consider the terminus of the implementation chain. The service should properly have been called Test, Trace and Isolate, for a significant aim was to ensure that known contacts of positive cases should themselves isolate. In May 2020, SAGE advised that a high level of public adherence to requests to self-isolate was required for the system to be effective. The likely rate of adherence was, of course, unknown but properly a vital part of the delivery model. Identifying potential weak spots is the raison d’etre of ex-ante evaluation.

NOA (2020) report that preliminary surveys ‘conducted between March and August indicated that around one in five respondents who had symptoms of COVID-19 fully self-isolated, and that one in 10 respondents who had been notified they were a close contact of somebody testing positive had isolated for 14 days’. These woeful indications led to the introduction of a supplementary system of fines and support payments – sub-interventions, whose effectiveness is as yet unknown and threatened by a welter of other concurrent innovations such as the fluctuating 3 Tier system of control and the initial roll out of the vaccination.

This particular episode is a microcosm of the Test and Trace programme. It is confronted in that famous phrase by ‘known knowns, known unknows and unknown unknowns’. Some unanticipated problems are dealt with, quite properly, by ongoing adjustments and supplements. As noted, all interventions have an element of muddling through. However, the NOA report and the work of independent journalists gives a clear indication that a more thorough ex-ante evaluation would have left a £20 billion scheme better prepared to anticipate what might be achieved and what might not.

2 Organisational learning. Another group of evaluation methods, particularly pertinent in the face of rapid changes needed to combat the virus, comes under the banner of ‘organisational learning’. Again, we encounter a family of approaches, known also as ‘continuous improvement’, ‘quality circles’, ‘developmental evaluation’ and so on. The approach has found usage in the management of all types of organisation but is particularly associated with healthcare settings that involve complex, resource-intensive treatments (Chrusch et al, 2016). The basic methodology is described in Figure 2.

2 The NHST&T script is not available publicly. Another version with a typical format can be viewed at: https://www.astho.org/COVID-19/Making-Contact/Scripts-for-Contacts/
The effective delivery of such treatments always faces technical, demand, staffing, safety, cost and other hurdles. Performance, moreover, always varies between the many responsible departments and units. Practice in relation to particularly pressing issues is monitored and reported in a sample of institutions (inner circles). Comparisons are then made in a ‘conference’ or ‘review’ (outer circle) in which potential solutions, best practice and benchmarks are identified. Ideally this sequence in continuous; underlying conditions often change requiring a ‘reset’ of approved approaches and standards. The harmonisation of this research sequence is difficult. The goal is not to supply regimented solutions but contextually sensitive options.

Figure 2: Organisational Learning

The value of organisational learning during the pandemic is considerable, particularly in relation to intensive care units (ICUs) which bear the brunt of care in the most severe cases. Models produced at the start of the outbreak provided dire and woefully inaccurate projections about how ICU capacity would be exceeded (for details see Pawson, 2020). On every upturn in the policy response, on every tightening of lockdown, grave declarations are made about their importance in protecting ICUs from being ‘overwhelmed’. Such prognostications are made on the assumption that ICU is a passive incumbent of the virus rather than an active, learning organisation.

ICUs did not (and will not) put up the ‘house full’ sign during the epidemic because they adapted quickly employing a rapid version of continuous improvement process depicted in Figure 2. Armed with knowledge of those developments, the senior UK medical colleges joined force to hurl a furious response to a story in the Sunday Times claiming that clinicians had drafted a triage tool to deny ICU care to frail and elderly (HNS England, 2020).

The reconfiguration of ICU began on a unit-by-unit basis. It started, literally, by learning from mistakes. Faced with a virus with unknown properties and few existing resources to contain it, failures became evident and demanded an immediate response. Belonging to a tight professional community, ICU specialists began to blog about how they adapted to and overcame the initial surge in critical care needs. More formal research interest immediately picked up and by April 6 The Lancet published a paper by Phua et al (2020), which gathered together an extensive record of how ICU practitioners and administrators had transformed critical care. Recommendations were offered on how to improve: access and triage; treatment variation, infection protection in the unit; the donning and doffing of PPE; the ventilation of units; the spacing of beds; logistical planning for equipment; consumables and pharmaceuticals; workforce load and augmentation; workforce communication and support; patient medication; collecting respiratory tract samples; post-ICU care, etc.

In the following months, this advice multiplied and became formalised and professionally endorsed. The Faculty of Intensive Care Medicine (2020) produced guidance on adaptations to critical care

Two short, final comments on organisational learning are pertinent. Whilst critical care services transformed to deal with COVID-19 at the sharp end, the virus drove also changes across the entire spectrum of healthcare. These resulted in so-called resets in all provision, which were also suggested and digested using continuous improvement cycles (Health Foundation, 2020). Primary care services, for instance, introduced digital triage, remote consultation, safer examination procedures, and new ‘step-down from hospital’ services. These are considered ‘resets’ because many of them solved existing demand management problems and are thus expected to become the ‘new normal’.

Medicine, of course has always aspired to be evidence based and the examples above all speak of a service familiar with improvement methodology. There is no organisation in the world that would not benefit from similar scrutiny – but alas few are equipped to do so. For example, Social Care in the UK is chronically underfunded and without a research infrastructure. There were outstanding examples of individual homes transforming safety during the pandemic – but as a sector widespread organisational learning came too little and too late.

3. **Community resilience research.** The success of governmental guidance on controlling coronavirus depends on its public reception. Only if the great majority of people adhere to the recommendations is the spread of the disease attenuated. Just as the virus is selective in its prevalence and impact there are wide differences in the ability and willingness of different communities to heed that advice. This suggests that another body of research is needed that attends to the micro-circuits of disease at the community level – rather than the default assumption that national or tiered regional interventions contain the solution.

The notion of community reliance in confronting national crises is not a new one (South et al 2020). For many a year the World Health Organisation has urged that building responsive and reliant communities is a key public health priority (WHO, 2012). Public Health England published guidance on a whole-systems approach to community-centred public health in the weeks before the virus arrived in the UK (PHE, 2020). On this model communities are not mere recipients of policies; rather they initiate and design them. The key actors are neighbourhood self-help networks, voluntary and charitable organisations, and city-level housing, care and health services. There are hundreds of variants of these bottom-up interventions, vital to the overall struggle against the virus but largely ignored in the political shouting-match on lockdown rules, with the associated research funded for peanuts when compared to the costs of the national inquires discussed earlier.

In the COVID era, there has been a massive growth of such services providing emergency payments, food aid, translations of advice and guidance, collection and delivery services, good neighbour packs, remote befriending, mental health support, help with children’s education and emotional support,

³ Note that ‘continuous improvement’ never equates to ‘definitive solutions’. Despite the incessant efforts at surge planning in ICUs, new problems continued to emerge, particularly in relation to the specialised nursing workforce and the inadequate time available to train additional recruits.
etc. The specific services provided are perhaps of less pertinent than their local know-how and network structures. *FoodWise Leeds* (2020), for instance, has buttressed the existing local system of foodbanks, soup kitchens and church provision, with new contributions from local business, council grants, charities, community events and a large donation network to become a direct provider to thousands of homes. Black, Asian and minority ethnic (BAME) organisations have responded to the crisis in a range of ways – advice in the right language, tailored messaging, information about rights, on-line community drop-ins, etc. *Sharing Voices Bradford* (SVB, 2000), for instance, has taken its mental health advice and counselling sessions online and now delivers community training on culturally competent support. In mosques there is detailed advice on the suspension of services, cleaning schedules, the removal of communal towels and prayer mats, burial for COVID victims, and so on (British Islamic Medical Association, 2020).

But this paper is about evidence not policy advocacy. I thus return to the vital question on the contribution of research – in this instance on community resilience. Little more needs to be said about its ‘Cinderella status’ in respect of national profile and funding. But this does not absolve it from questions of reliability and validity. Two short remarks are worthwhile. The first is that surveillance on the localised spread and impact of the virus is relatively well resourced, rigorous and thus better able to pinpoint need. An example is the West Yorkshire and Harrogate Intelligence and Insight Pack (2020). This contains a fine-grained examination of the impact of the virus on BAME communities (most notably Bradford, with its high proportion of Asian British residents). Prevalence, mortality, and shielding rates are pinpointed minutely at the ‘ward’ level. Perhaps more significantly, material is collected on what the report calls ‘soft intelligence’. This identifies why specific communities have fared badly under the virus – e.g., collapse of the local ‘cash-in-hand’ economy, significant exposure to ‘fake news’ media, cultural misunderstandings with providers and referral systems, stigma involved in using city-wide services, inter-generational conflict in households, social distancing problems with large families in small houses, curtailment of funeral and mourning services, and so on. Detailed knowledge of the problem then arms the solution.

This brings us to research evaluating the effectiveness of the array of community resilience schemes. Again, there is no immunity from the classic methodological questions on outcomes and effectiveness. What do we know about whether they work? Here, I reach the opposite conclusion. By and large these voluntarily funded, locally managed projects are poorly evaluated. There is, of course, a sound reason for this – finance is tight, so delivery rather than evaluation becomes the priority. The paucity of evaluation can be seen directly in the COVID projects listed in the research repository of the Voluntary Sector Studies Network (2020). In a manner that is entirely typical of these underfunded, small-scale studies, the research reports are largely descriptive and depend for evidence on the testimony of staff and participants. Unsurprisingly, they read as ‘good news’ stories.

We know, because it is true of all intervention, that community resilience programmes will only work for some residents, in some contexts, in some respects. This eventuality has never been put more frankly than in the aforementioned British Islamic Medical Association Guidance (2020):

‘Please note: In situations such as this, there will always be a wide spectrum of opinions on what are appropriate precautions to take given the circumstances, so consulting your local scholars early is essential ... We would advise everyone – especially those who may disagree with some of the above guidance – to weigh up their position against the impact in what may be a life-or-death situation, especially for the most vulnerable in our community’.

I draw a different, methodological lesson on the challenge of evaluating the ‘spectrum of opinion’ on community programmes. These interventions tackle the consequences of national policy at the
sharp end. These citizens, along with key workers and care home residents, bear the brunt of the virus. The case for more community action and more local resources has been made many times (Rippon et al, 2020). But my point is that the ‘the science’ belongs here too. Quelling the virus depend on understanding and then controlling its localised spread. Much is understood about vulnerable localities, yet we know little about what works and what doesn’t work within these micro-circuits. There is a clear demand for a programme of critical and systematic evaluation. The old cry that ‘more research is needed’ has never been more apt.

Conclusion

What came before coronavirus? Societal transformation in 2020 was so overwhelming that it is sometimes hard to recall. History will remind us, however, that it was the ‘post-truth’ era. ‘Lies and deception, exaggeration and euphemism, flannel and waffle, spin and the selective use of the facts, artifice and insincerity’ were everywhere (Davis 2017). With notable success, populist politicians leapt at the opportunity to confuse ‘facts’ and ‘alternative facts’. In the UK, a series of elections and a bitter referendum were won and lost with little care for the truth.

Thus, to a big concluding question – has politics and policymaking changed as a result of the pandemic? To be sure, some appetite for fakery continued into the depth of the virus. Former President Trump was content to lead the faithful by finger pointing at the ‘China virus’ and with his scholarly wisdom that disinfectant could slay COVID-19. Many social media sites continue to this day with bogus claims about immunity and the effectiveness of snake-oil treatments. But by and large politicians have reached out for scientific evidence. This decision-making process was made manifest in microcosm in the scores of televised bulletins that occurred in England in 2020. From the oak panelled state rooms of Downing Street, the Prime Minister flanked by the Chief Scientist and the Chief Medical Officer presented information on the vital signs of the epidemic, pronounced on gravity of the situation and on the urgency of the requisite response. The PM harrumphed with the usual grandiloquence, but always with enormous and genuine deference to the left and right about their scientific warrant.

So, has the right balance now been struck? Has evidence-based policy come to pass? I remain unconvinced and have made the case for more usage of complexity-informed policy analysis. A programme of organisational learning, of ex-ante process mapping, and of community-resilience research would have brought a better balance and more rigour to the evidence base. It would also have introduced a proper modesty about the fallible grasp of evidence. Politicians and policy makers, of course, are always unsettled by complexity. The same, of course, goes for the mainstream media. They much prefer ‘soundbite science’ and professorial pundits have not disappointed, queueing up to make self-confident pronouncements on how, when and with what severity that lockdown should be delivered.

Very rarely did they feel able to utter the authentic scientific response, ‘as yet, we don’t really know’.
References


West Yorkshire and Harrogate Health and Care Partnership (2020) Intelligence and Insight Pack https://www.wyhpartnership.co.uk/application/files/4515/9713/9130/Paper_B_Developing_an_Intelligence_and_Insight_Data_Pack.pdf
