HELEN MACNAMARA　　　　From:　　DAVID HALPERN
ALEX CHISHOLM　　　　　　cc　　　　NR

　　　　　　　　　　　　　　Date:　　28 September 2020
　　　　　　　　　　　　　　　　　　July 2020 (orig. draft)

INSTITUTIONAL LESSONS FROM COVID

You asked for some thoughts on the lessons from HMG handling of COVID. This note is a personal view, albeit situated in previous work on institutional weaknesses and failures, including *Behavioural Government*[1], King and Crewe's *The Blunders of our Governments,* and Eggers' *If We Can Put a Man on the Moon….* The summary points are:

1. **The early misstep.** Overconfidence and anchoring in our expert medical community led to a presumption that covid would be a flu-like wave, blinding it to the pursuit of near-suppression as a viable option and an expanded tracing system in particular. Our decision-making process was vulnerable to systematic error.
2. **Don't put your eggs in one basket**. Where daring punts were made on single solutions, such as the first NHS app, they generally failed, whereas when multiple solutions were tried, eventual success was achieved.
3. **'How?'** People mistook academic enquiry for policy. We neglected the 'engineering' of effective delivery.
4. **'Authoritative' versus 'authoritarian'.** The policy and comms strategy sought to occupy a position of 'soft authoritarianism', based on simple, rigid rules yet with weak enforcement. Arguably, we would have done better going for an 'authoritative' approach: risk and principles-based guidance that could be flexed.
5. **Putting humpty together.** A strong bio-medical perspective was not effectively tempered with robust behavioural or economic analysis, and a weak central process failed to give adequately balanced advice to the Cabinet of PM.

There will be another covid. Alongside improvements to core decision making, we should use civil service reform and the SR to strengthen the empirical and methodological competence of the UK policy profession. I am also going to seek to build a UK 'social and behavioural MIT' - an world-class institution that can deliver timely advice and solutions to the standard we need.

The puzzle

The central puzzle is how the UK brought together a world class group of experts, alongside a world class civil service, that together led to such an un-world-class

---

[1] https://www.bi.team/publications/behavioural-government/

outcome. The UK remains in the 'top-10' worst national performers by per capita death rates, despite having had the benefit of a lagged start. Many other countries have achieved 10-fold, or even 100-fold, lower per capita death rates than the UK (614)[2] - such as Norway (48), Australia (35), S Korea (7), NZ (5), Singapore (5). There are many factors involved in these differences, but we should urgently learn lessons, not least in anticipation of responding to the secondary social and economic shocks to come.

The 'easy' response is to blame the politicians. But the failure was much more extensive, and more subtle. Weaknesses in the UK response fit with James Reason's 'swiss cheese' account of failure[3]: an alignment between multiple institutional and human weaknesses, as seen in historic failures from 'Three Mile Island' to 'Deepwater'. It is also important to examine the counterfactual - countries or parts of our own system that performed better. Reason et al called these 'highly reliable' organisational forms: highly complex organisations, such as aircraft carriers or air traffic control systems, that are subject to multiple and varied challenges yet consistently do not fail.

1. The early misstep: overconfidence and anchoring

Arguably the most fundamental misstep in the UK response was the presumption that covid would be an unstoppable flu-like wave. This presumption was built into the Contain-Delay-Mitigate-Research strategy published in early March. It also underpinned the early (Chris Whitty) position on test and trace, and the Vallance view on 'herd immunity' (later air-brushed).

It is important to see that this presumption was not based on ignorance, but on a century of prior knowledge and assumptions. SPI-M had sophisticated models drawing on data going back to the 1918 flu showing the dynamics of prior epidemics. Chris Whitty had prior field experience, and was extremely concerned to get the 'wave' landed before winter. Jonathan Van Tam wrote the textbook on flu, and had strong views on the likely modes of transmission and spread. Patrick Vallance knew that vaccines would not be available till the end of the year at the earliest.

All this massed expertise converged on the conclusion that, once early containment had failed, a flu-like wave was inevitable. As such, the best that could be done was 'to flatten the curve'. This would take the worst of the negative effects out by (a) preventing the NHS being overwhelmed, (b) stopping the epidemic over-swinging (ie achieving 'herd immunity and around 60% instead of 90%), and (c) protecting the most vulnerable until the wave had passed.

---

[2]

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1117770/

But there was an alternative. The world - and more specifically the South Asians - had found another strategy: to build a sophisticated test and trace system strong enough to substantially suppress the virus, at least for long enough to enable treatment and vaccines to be developed. This is not just a matter of hindsight bias. There were a number of us - generally outside the biomedical bubble - who highlighted this alternative strategy. BIT studied the responses and techniques employed by S Korea, Singapore and others from the outset, including the particular techniques they used to do the tracing, and which of these might be transferable to the UK context (eg phone tracing and notifications, using commercially available tracking). This led us to question the limited ambition of the central 'delay' phase, though we found ourselves quietly dismissed as not really understanding the science.

If there was a single top lesson to be taken from the 2018 Behavioural Government report, it was 'beware over-confidence'. Humans are fundamentally prone to overconfidence in their own beliefs, systems and groups. This overconfidence tends to become more serious the more senior the individual.

Experts and academics are prone to the same weakness. Despite a century of statistical methods to directly estimate confidence, the academic world remains characterised by overstatement and overclaiming. Indeed, whole careers are built on battles between rival academics overstating their cases in the pages of academic journals, with corresponding overestimates of effect sizes from publication bias.[4]

There were multiple examples of such 'anchoring' within SAGE, with early hypotheses and views stuck with, despite mounting evidence against. Early evidence of low german death rates was repeatedly dismissed as being the result of cases occurring in 'fit skiers', and that their death rates would shortly converge on the UK numbers. The role of aerosol transmission and corresponding importance of ventilation was underplayed (relative to ballistic). Asymptomatic cases, and their policy implications, were understated. The growing evidence on masks was very slow to be taken on board.

At the same time, a proposal (supported by Matt H) to create an open wiki-style list of questions, or known unknowns, was quashed by PV. Instead PV asked Mark Walport at UKRI was asked to organise a website publishing existing research: a (delayed) shop-window of what we already knew, rather than a pinpointing of what we didn't. Ironically, the pride in our science and our capabilities, slowed our ability to learn lessons from other countries. Under cover of variations of 'it is very different there', there was an arrogance that we knew better, and would do better.

If I had to choose a single issue - and moment - that embodied this failure, I'd point to the unshakable conviction of the SPI-M modellers that suppression - the sustained holding down of covid prevalence - was not a viable strategy. For example, Graham

---

[4] https://eml.berkeley.edu/~sdellavi/wp/NudgeToScale2020-05-09.pdf

Medley, chair of SPI-M, was asked in SAGE in the second week of March, along with the other modellers: 'how certain they were that major second waves would arise in China and other Asian nations?' (ie that suppression was not a viable strategy). Medley answered immediately, and with total conviction: 'as close to 100% as possible...'yes, 100%'. His colleagues echoed the conclusion 'yes, 100%'. They were totally convinced that as soon as the harsh lock-downs in Wuhan, South Korea and elsewhere were lifted, cases would immediately surge again.

Nothing in science, and certainly in statistical modelling, is 100%. Let alone in the face of the data already emerging from the Asian experience. This was doctrine, not science.

Don't: get carried away with your own models (and ego)

Don't: shut down counter or heretical views, or seek an overly early 'single view of the truth'. Avoid having key meetings chaired by people with strong prior expertise in a particular area or specialism.

Do: use scenarios that ensure your policy is robust to a range of possibilities, and/or supplement with 'red-teaming' to challenge strong priors

Do: have key meetings chaired by intelligent generalists; seek to develop alternative viewpoints, and use mechanisms to keep questions on the table.

2. Don't put your eggs in one basket: diversity and innovability

Anchoring and rigidity was not limited to our core epidemiological assumptions. It also dogged our delivery. We generally succeeded when we set higher level functional goals and challenged innovators to deliver. We generally failed when we over-specified the objective and means to deliver it, and then bet the bank on a single solution.

The decision by NHSx to put all its eggs in one NHS App basket was a particularly grim example. When the idea was proposed for an app (March 6th) we, along with Ben Warner from No10, strongly supported it being taken forward, but also strongly recommended that this be done as part of a portfolio. We argued that more than one approach to the app should be developed (such as setting up 2 or three parallel teams, at least up to an early gateway); a re-skin of an existing app be attempted (eg Singapore, Australia); and non-app based approaches should also be pursued (eg SQREAM). NHSx not only pursued only one (internal) option, they persisted with this as deadline after deadline passed. This was a serious policy failure in its own right. A version of a tech-enabled app should, and could, have been delivered to be ready for the end of the first lockdown (when its value really would have kicked in). The app, delivered by a separate team after the first was shut, was delivered six months after the original target date.

4