

1: A conversation on a train: Why regulate official statistics?

A couple of years ago, back when it was possible to travel freely without worrying about masks, infections and tests, I was on a train journey south from Edinburgh.

As is the way with long train journeys in the UK, there was some horrendous disruption; and as is also the way, this disruption broke the invisible veil that holds British people back from talking to one another, and lots of conversations started – proceeding from the usual starting point of “bloody typical” to broader chats – where are you going, what do you do?

I got chatting to one person in particular, who was heading for the final part of an assessment process to be admitted to the United Reformed Church as a Minister. So we spoke about that for a while, and I learned one thing I’m very happy to pass on: don’t ever apply to be a Minister in the United Reformed Church: it would be less arduous and time consuming to go through the SAS training programme.

Anyway, eventually, the conversation turned to me, and what I do for a living.

And instead of me needing to explain in depth and at great length the role of statistics, what they are, how Government can use them, how Government can misuse them; instead of me having to justify my purpose against an implicit question “is that really a real job?” – instead, in short, of the conversation of bafflement and perplexity that I feared, this proto Minister immediately got it; understood what we do; and why it was important.

And in fact, they put it in stunningly simple terms. What they actually said was “I get it, what you do is....”

...well, I’ll tell you what they said at the end of the talk.

Because this afternoon I want to focus on this issue of what we do as statistics regulator and why it is important. In fact, I’m going to focus mostly on the “why”; I’m going to start there,. Why do we regulate statistics? As I’ll come on to explain, it’s because they represent a public asset, part of the fabric of democracy; and because they can be misused by the people responsible for producing them.

I’ll focus on the ‘why’ for three reasons. First, because it’s the most important part of any organisation’s mission. Second, because while some of the detail of what we do is quite specific to our legislation, the ‘why’ is more accessible, more universal, and speaks to some fundamental features of contemporary society and democracy.

And third, because the idea of statistics regulation may appear to stretch notions of what regulation is for. I want to emphasise instead why regulation of statistics is important, in fact essential, and how our purpose is very much in line with the way the RPI as an Institute thinks about regulation.

I’m going to start with regulatory philosophy. Then I’m going to outline briefly what we do at the Office for Statistics Regulation, using three recent examples of our work. Then I will explore the “why” – this notion of statistics as a public asset that can be harmed by misuse. I’ll move on briefly to the “how” – how we implement our regulatory approach, and closing with a summary that considers the alignment of our approach and the RPI view of the world.

The RPI world view

On the face of it, there is a difference between the oversight and regulation of official statistics on the one hand and the economic regulation of sectors, like energy and water. Official statistics are not traded in a market, with consumers paying for services from providers. They are not transmitted down expensive physical infrastructures. Official statistics may not seem as essential to life as energy, water or telecommunications.

RPI is focused on promoting public understanding – looking beyond the technicalities to the underlying dynamics. This creates a strong tendency within RPI – I say tendency, in preference to philosophy or anything that smacks of a single Received Wisdom

Here is what I regard as four of the key features of the RPI tendency:

1. *Regulation exists to prevent harm.* That harm is to the interests of consumers – who can be over-charged or who can receive poor services.
2. This harm arises because consumers are at the wrong end of *the asymmetry of power and information*. Their interests are perpetually at risk of being side-lined in favour of the interests of the large, asset-rich businesses that provide their water, energy and so on; and consumers are also at risk of being disregarded by policy makers in Government.
3. It's summed up neatly in the concept of *abuse of dominance*. Regulation exists to balance up these asymmetries; to ensure that these large powerful interests do not run things in a way that suits them.
4. It follows from this that regulation must be, and be seen to be, *an independent force* – distinct from commercial and political interests.

There is another strain of RPI thought that I always find attractive – which is to mistrust 'inadequately examined' simple answers. Simple answers can sometimes be found, but discovering the best answer (whether simple or complex) always requires engagement with the complexities of the context (due consideration/examination).

And underlying this is a respect for the foundational thinkers in economics – for example, the Hayekian idea that markets are superior to centralised planning because they process information much more effectively through the decisions of masses of individuals; and Adam Smith – not just the famous invisible hand, but also Smith as a moral philosopher. I often feel that RPI regards regulation as much of a social as technical endeavour. And when I shared this speech with George, he shared this summary of Smith's philosophy, taken from a past edition of the Stanford Encyclopedia of Philosophy (under the entry 'Adam Smith's Moral and Political Philosophy'):

“A central thread running through his work is an unusually strong commitment to the soundness of the ordinary human being's judgments, and a concern to fend off attempts, by philosophers and policy-makers, to replace those judgments with the supposedly better “systems” invented by intellectuals. In his “History of Astronomy”, he characterizes philosophy as a discipline that attempts to connect and regularize the data of everyday experience; in TMS [the Theory of Moral Sentiments], he tries to develop moral theory out of ordinary moral judgments, rather than beginning from a philosophical vantage point above those judgments; and a central polemic of WoN is directed against the notion that

government officials need to guide the economic decisions of ordinary people. Perhaps taking a cue from David Hume’s skepticism about the capacity of philosophy to replace the judgments of common life, Smith is suspicious of philosophy as conducted from a foundationalist standpoint, outside the modes of thought and practice it examines. Instead, he maps common life from within, correcting it where necessary with its own tools rather than trying either to justify or to criticize it from an external standpoint. He aims indeed to break down the distinction between theoretical and ordinary thought. This intellectual project is not unconnected with his political interest in guaranteeing to ordinary individuals the “natural liberty” to act in accordance with their own judgments.”

This fits with my contention that we should regard regulation as a social, not a technical, endeavour, and it infuses the rest of this lecture.

The OSR world view

Let’s step away from this RPI world view, and instead talk about what we do at OSR, what we worry about, and the risks and benefits we see. Let me give 3 examples.

In May last year, Matt Hancock, then Secretary of State for Health, was fond of saying that the Government had set a target of 100,000 tests a day. And every day, a figure was published that showed how close or far from this target they’d got.

We stepped in publicly to challenge this “100,000 test” measure. That’s because it wasn’t clear what the definition of ‘test’ was, and whether it would align with what a member of the public would think of as a test. A ‘test’ was counted as taken if a testing kit was dispatched in the post, not when it was actually used, returned and the samples processed. It was also not clear what actually counted as a “test”. The notes to the daily slides said that some people may be tested more than once and it was even reported that swabs carried out simultaneously on a single patient are counted as multiple tests. But it was not clear from the published data how often that was the case, if at all.

So we wrote to Matt Hancock and said that “The aim seems to be to show the largest possible number of tests, even at the expense of understanding¹” and that therefore it was hardly surprising that people didn’t have confidence in the figures.

As a result, the Department for Health and Social Care dramatically improved the rigour and transparency of how it reported tests, and we have published supportive statements around statistics – including a couple of ‘rapid review’ assessments².

A second example surrounds the daily news briefings. On October 31 2020, the Government announced a four week lockdown. The briefing was supported by a set of slides delivered by the Government’s chief scientific advisors. And after looking at the presentation, and the slides, we wrote to those advisors – Sir Patrick Vallance and Sir Chris Witty – to highlight a

¹ <https://uksa.statisticsauthority.gov.uk/correspondence/sir-david-norgrove-response-to-matt-hancock-regarding-the-governments-covid-19-testing-data/>

² <https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-lucy-vickers-review-of-nhs-test-and-trace-england-and-nhs-covid-19-app-statistics/>; <https://osr.statisticsauthority.gov.uk/wp-content/uploads/2020/07/Ed-Humpherson-to-Stephen-Balchin.pdf>

series of problems with the material³. The core of our concern was transparency – the data underpinning the slides was not available publicly at the time of the media briefing, and some of it – like the worst case scenario – was very opaque. We judged that it was unreasonable to expect the public to take on board such a significant change to their lives if the evidence was held so closely within Government.

Our letter to these advisors became one of our most viewed web pages. To their credit, Patrick and Chris acknowledged the issues⁴; the data were made much more available; the briefings have improved hugely thanks to the support of ONS; and the principle (if not always the practice) of transparent publication of data is much more firmly established.

My third example is the exam grading⁵. You'll recall the story. Exams were cancelled for the first time in 2020, and instead the grades for students for A-level and GCSE were set by a statistical model, loosely called an algorithm in popular discussion. When the results came out, some people didn't get the grade they were expecting; there was a public outcry; and the model was abandoned. Instead, the students were awarded the grade predicted by their teachers.

We stepped in again. Our concern was that people felt that statistical models had been used in a way that damaged their interests – and not only was that significant in its own right, but also called into question broader public confidence in statistics. That struck us as important because we expect the use of such models to increase in public life in the coming years.

We undertook an extensive review, across the UK, and published our report in March last year⁶. Our report avoided the superficially attractive approach of lambasting Ofqual and the equivalent bodies in Scotland, Wales and Northern Ireland. In fact, we commended them for being dedicated public servants trying to resolve an incredibly difficult problem. But we did highlight some core problems with the approach – most of which were not technical in nature: this is not a story of rogue algorithms; but instead of public confidence. The key point is that we used our report to set out a series of things that any public body using models should consider – and they revolve around transparency, engagement and quality assurance.

That then gives a flavour of what we do.

Seduction of data, data everywhere

Why do we do it?

We live in a world seduced by the power of data. People talk about data being the new oil.

Of course, this an imperfect analogy. You use oil up, but you don't exhaust data in the same way. And using oil creates some nasty environmental externalities, which isn't the case with data. It does have one neat feature, though: raw oil is fairly useless; it needs to be processed into something else to be useful. So it is with data: the useful processed product is statistics.

³ <https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-sir-patrick-vallance-transparency-of-data-related-to-covid-19/>

⁴ <https://osr.statisticsauthority.gov.uk/correspondence/sir-patrick-vallance-to-ed-humpherson-transparency-of-data-and-information-related-to-covid-19/>

⁶ <https://osr.statisticsauthority.gov.uk/publication/ensuring-statistical-models-command-public-confidence/>

Anyway, the weakness of the analogy hasn't stopped lots of people being seduced by the power of data, including in Government. There are multiple data strategies; and multiple examples of Government wanting to demonstrate that it is data savvy, data led, and, in a pandemic twist, following the science.

You see it too in the civil service's ambitions for the future. In the Declaration on Government reform⁷, from June 2021, there is a feverish excitement for data: we read that Government must do "better at pooling and sharing our data"; better at "making our data available"; developing expertise in "digital, data, science"; that "data when used well is telling us so much more about what works and what does not".

But this almost utopian narrative of the power of data is accompanied by a degree of anxiety. One source of this anxiety is simply volume – that there is so much of the damn stuff. A quick google of the term "data, data everywhere" reveals that this anxiety – that we have so much data but we don't know what to do with it -is something of a meme. Data scientists appear to enjoy involving "data data everywhere but not a drop to [insert verb: drink/use/generate insight from]". Perhaps they like to show that they are more than just data scientists, that they have an extensive literary hinterland. In a delightful irony, of course they are all getting the quote wrong when they render it "data data everywhere but not a drop to use". Look it up to see the phrasing of the original poem.

Data abundance also invokes Gresham's law. This holds that when a currency is debased, the bad money, the false coinage, drives out the good⁸. It is often used as an analogy – to a situation where people lose confidence in the value of something. For data, the analogy goes like this: the bad data – the fake news, the misused statistics – drives out the good. People no longer know what to have confidence in. They mistrust everything. They see the world of quantification and evidence is inherently untrustworthy; they cash in their chips; they think it's all made up.

As an aside Gresham never actually used the phrase "the bad money drives out the good". This was a reinterpretation of his writing by a much later economist

Is there any evidence of this happening? Well, it's certainly possible to find evidence of the debasing of the currency of data and evidence. There are famous, iconic examples – insert your own favourite misuse here.

There is evidence too in the nature of discourse. A couple of years ago my wife got a book called Great Speeches of the 20th century. these Great Speeches – all the *fighting on the beaches*, the *I have a dream*, the *ask not what your country can do for you* – don't feature many numbers, are not awash with data or claims about data. Compare that with a speech of today. In day-to-day politics – Parliamentary question times, for example – you see a bombardment of numbers across the despatch box.

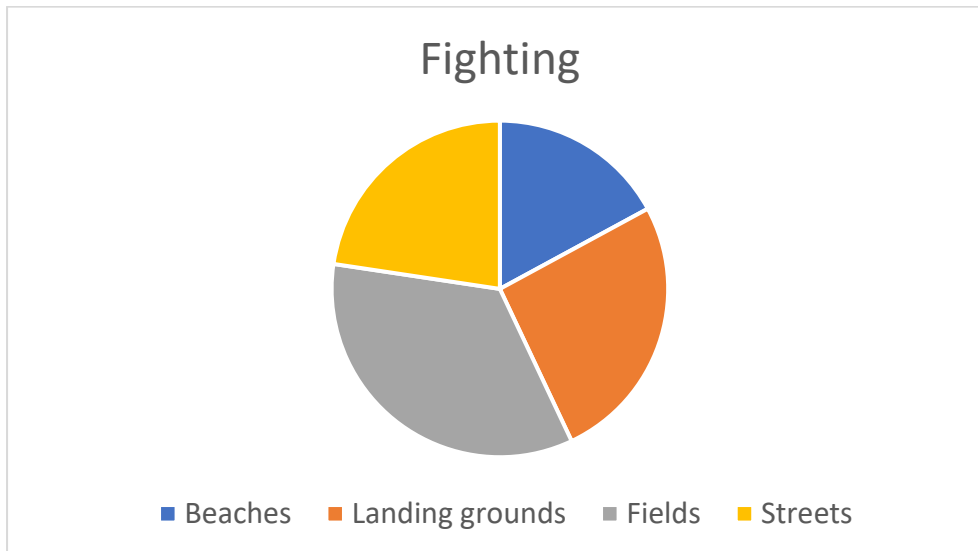
⁷https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993902/FINAL_Declaration_on_Government_Reform.pdf

⁸ Thomas Gresham. He was a Tudor banker, initially based in the Low Countries, but ended up working for the royal household. He is known for his concerns about debasing of a currency. "Debasing" a currency is when the composition of a coin is deliberately changed – in other words, precious metal is replaced by base metal, gold replaced by an alloy. Gresham's Law states that the "bad money drives out the good": confidence collapses, people don't know which money to trust, and they stop using the money altogether.

It made me wonder what a Great Speech of the 20th century would like now. I think Churchill's beaches speech might go something like this:

I can today announce we shall fight on the beaches 17.3% of the time, we shall fight on the landing grounds 26.2% of the time, we shall fight in the fields 34.8% of the time and thanks to the Government's long term plan, I can today announce that we will fight in the streets for a further 21.8% of the time.

Perhaps accompanied by a pretty but meaningless visualization:



(And in the interests of provenance, let me just emphasise that this is not a real Churchill quote).

There's also a broader sense of a threat to public reason. Baroness Onora O'Neill covered this in a lecture on Ethical Communication in a Digital Age a couple of years ago⁹.

By public reason, she meant the notion that public life should be based on and protect certain standards and norms. She outlined how digital technologies undermine these standards and norms. Information and communication technologies are not used to inform or communicate, but to grab attention, and use it to misinform, nudge and manipulate.

She also made a fascinating observation. We tend to spend more time thinking about the rules that support public reason than about the ethics. In other words we focus on freedom of expression, rule of law, separation of power, legal obligations on privacy. We don't think much about things like veracity, responsible use of evidence, integrity – indeed they are seen as somewhat quaint. In a neat summation, she says we spend more time on what it means to be public than what it means to reason.

Why does this matter? In our work as the statistics regulator, we see statistics as an asset. Statistics start with data and create insight out of those data: addressing quality issues, describing what the data mean. Statistics come with a credibility and assurance of professional review that you do not get with data.

Statistics frame our understanding of the world. They help give a starting point for debate – on the size of the economy, the number of people in the country, the rate of crime, the health and well-being of the population, and many other things. They support choices.

⁹ <https://www.thebritishacademy.ac.uk/events/ethical-communication-digital-age/>

Choices by institutional decision makers – like the Bank of England, or a Secretary of State. But statistics aren't just for these official purposes. They also support the choices made by a very wide range of people individuals, businesses, community groups and so on. And statistics can be passively consumed, incorporated almost imperceptibly into choices – like the sense that people have that there may be long waits in A and E, or that crime may be rising. This doesn't come from actively studied statistics, but passively consumed statistics – things heard on the radio, through the media.

And of course, statistics are a crucial part of democratic accountability – enabling people to hold policymakers to account for the impact of the decisions they've taken.

So these things, this risk of debasing, represent a series of threats to statistics as a public asset.

Our regulatory role

So my starting point is the notion of statistics as a public asset. With a wide range of uses and users. And which can be harmed by misuse, overuse, and by a general debasing of the coinage of good data.

In our view, there is a right to have access to high quality information and data that is publicly available – it is why statistics are valuable for our democracy and democratic accountability.

Furthermore, there are producer interests at play here. For every statistic that you hear, there is a producer. By producer I mean the professionals responsible for collecting, compiling and publishing a set of statistics, and USING THOSE STATISTICS, within an official public body. So “producers” refers to the Office for National Statistics. But also Government Departments, agencies, devolved administrations, and their spokespeople – the people who use statistics to communicate key messages. This is what I mean by “producer”.

I said earlier that, on the face of it, there is a difference between the oversight and regulation of official statistics on the one hand and the economic regulation of sectors, like energy and water. Official statistics are not traded in a market, with consumers paying for services from providers. They are not transmitted down expensive physical infrastructures. Official statistics may not seem as essential to contemporary life as energy, water or telecommunications.

But what I hope is starting to emerge is real similarities. Official statistics are seen as a fundamental building block of democracy – for example by the UN Fundamental Principles of Official Statistics¹⁰ – and in that way as much as an essential feature of life as traditional utilities. Official statistics have public good characteristics. And while government Departments may not have a complete monopoly – private organisations can collect data and aggregate them into statistics – there are reasons to regard official statistics producers as facing the same risks of pursuing producer interest over the consumer interest as have long concerned regulators of utilities.

In the statistics context, this *producer interest* can be redefined as the vested political interest of the Government of the day, and the *consumer interest* as the public good that is

¹⁰ <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

served by reliable, freely available statistics. One of the main objects of statistics legislation and regulation is to protect official statistics from political interference.

Trustworthiness, Quality and Value as a sociotechnical approach

How then do we regulate?

Basically, we set standards for the production and publication of statistics. And these go beyond being a technical set of criteria for how to add things up. They are in fact based on a sociotechnical approach – that idea that technologies (like statistics) only have meaning and power in relation to society and people’s reaction to them.

We bring this together into a philosophy we call TQV: Trustworthiness, Quality and Value.

For statistics to do their job – to serve the public good – it’s not enough that they are collected well and adhere to reasonable standards of accuracy, timeliness and reliability. They must demonstrate three distinct features:

- Trustworthiness – they must be produced in a way that is free from vested interest. We see trustworthiness as a set of commitments by producers to act in certain ways that are not in their short term interests. For example, to commit to publishing statistics on a given time and day – and not therefore to allow the political exigencies of the day to influence the timing of release. Our Code of Practice¹¹ has a series of such commitments.
- Quality – this is about the statistics themselves: where they are from and how they’re collected; the approach to quality assurance. It starts from recognising that no statistic is an objective measurement of a hard physical reality – it is an estimate of something out there in the world; and as an estimate it is important to understand and test the basis on which it is made. And crucial is conveying the quality to users – so that people understand what the data are and their limitations.
- Value – ok, you can produce statistics in a trustworthy way, and you can be clear about the estimating basis. But so what? Statistics do not exist for their own purposes. They exist to help people understand the world, to answer their questions – to use a current example, to enable them to make sense of how much infection there is circulating in their community and to make choices about how they behave accordingly. We call this value – the extent to which statistics meet people’s needs, and answer their questions, and it’s underpinned by equal access to the information for all, so anyone can access the same information as seen by policymakers.

I won’t go into the detail of how we enforce these principles.

The beauty of this approach to regulation is that it does not focus on identifying and correcting negatives. Instead in OSR we have moved to a model where we also protect and encourage positive developments, championing producers where we can (more than ever during the pandemic). TQV – as we call it – encodes aspirations. After all, who wouldn’t want to produce statistics and data that are trustworthy, high quality and high value?

¹¹ <https://code.statisticsauthority.gov.uk/>

Furthermore, we are small. We can't cover everything – which is why we want the TQV aspiration to do so much work for us.

In addition to this aspirational incentive, we do a range of reviews and publish them; and step in when we see these ideas of TQV being infringed. let's go back to my examples:

- The 100,000 tests is really a failure of quality. It just wasn't clear on what basis the stated estimate of tests per day was being made. So we stepped in to demand it was made clear.
- The 31 October media briefing was really a problem of value. How could people make sense of this new lockdown, how could their questions be answered, if there was no equality of access?
- And the exams was as much an issue of trustworthiness (though both quality and value were also crucial here). Because essentially the outcry was based on a fear that the exam models had not been produced with the best interests of the students at heart, and the exam organisations were not able to provide evidence – of things like quality assurance and appeals – to convey their trustworthiness.

Closing remarks

Data is evolving and becoming more pervasive, and we are too – the exams work shows us focusing on algorithms, as opposed to the production of statistics themselves.

Is this really so different from what economic regulators do? I don't think so. Indeed, I'd say that the underpinning RPI tendency I outlined earlier is deeply intertwined with our approach at OSR.

We challenge abuse of dominance, the dominance over information and how statistics are used, their strengths and limitations and what data and statistics are even available. To put it another way, we challenge the asymmetry of power whereby a Government producer of statistics has all the power to collect, compile and present the data, and the consumer – the citizen, or business or charity – has none. That is consistent across all three of my examples – the 100,000 tests; the media briefing in October; and the exams.

In doing so, we are focused on preventing harms, just like any more traditional regulator.

And we have an inherent, almost Hayekian, mistrust of centralised decision making, which is why we put so much emphasis on this sociotechnical approach of TQV – confidence in statistics depends not on producers acting independently but in their engagement dialogue and interaction with the people who use statistics. It is inherently about the engagement between the owners and producers of statistics; and indeed one of the greater dangers is where producers allow their technical interests to dominate; or even, as with 100,000 tests, there's a sort of pseudo-technicality, a dressing up in technical-sounding language to try to pass something off as having the hardness and rigour of science when in fact, as with all statistics, it is a process of social engagement; back to Churchill and my imagined pseudo quantification of the beaches.

In RPI terms, there is:

- A risk of abuse of dominant position

- An asymmetry of power, with producers owning the data and choosing how and when to release them, and in what format and interpretation
- A strong sense that a static, centralised approach is inferior to one which engages with the users of statistics – something, that is, close to the Hayekian distrust of centralised decision making.

And there's one final point I'd like to make as an aside. It seems to me that economic regulation has become a highly technical endeavour, and I'm not sure that this is always the best approach. Is there something to learn from a sociotechnical approach, like TQV, one which places the technical in a broader framework, and where trustworthiness and value are as important as RAB and WACC? That's not for me to say, but perhaps worth the economic regulation community thinking about.

Conclusion

So this then is what we do. We uphold TQV. We aim to exert a countervailing power to the centralised interests of data producers. And seek to prevent harm.

But I'm not going to have the final word on this. This lecture has had as a recurring motif the way that quotes can lose their reference points and in the process become mangled. It's been a lecture peppered with things like the misquoting of the ancient mariner, the butchering of the beaches speech, the imprecision of the oil metaphor, the non-quote of Gresham's Law.

So I'd like to close with one verbatim quote, in its context, and retaining intact all its meaning.

Remember my train journey – when I explained what I do to the candidate for the United Reformed Church (and by the way, I never found out if she got the job). Let's give her the last word, because she immediately grasped what we do, and summed it up like this:

"I get it. You make sure that the Government isn't lying to us and telling us that things are better than they are.