The Strategy Unit.

# 'Rules of thumb' for decision makers approaching analytical work

# Document control

| | |
|---|---|
| **Document Title** | 'Rules of thumb' for decision makers approaching analytical work |
| **Prepared by** | Fraser Battye |
| **Checked by** | Peter Spilsbury, Fraser Battye |
| **Date** | 3rd April 2025 |

# Contents

# 1. Introduction

This report was produced by the Strategy Unit for the Health Foundation. It asks whether there are any rough 'rules of thumb' that decision makers (who are typically not analysts) could use to make broad, initial assessments of quantitative analytical work. The work is exploratory, designed to support analyst-decision maker relationships.

The quote to the right, from Amanda Pritchard - ex-Chief Executive Officer of NHS England - points to the value of analysts and analysis.

*"A good analyst can save more lives than a good anaesthetist."*

Pritchard suggests there are causal routes between 'good analysis' and 'lives saved'. She also suggests that this takes place on a greater scale than for anaesthetists. This view will be welcomed by many analysts. And yet there is an implied downside: that poor analytical work could cause harm on a scale that a poor anaesthetist does not.

Many have noted that the NHS is analytically underpowered (e.g. see the Goldacre Review). Despite the emergence of roles such as Chief Analytical Officer, senior NHS decision makers often lack access to high-quality and trusted analytical advice.

There is no substitute for this advice. To make reasonable and defensible choices, decision makers need support from high-grade analysts. Promoting benefit and avoiding harm depends upon it. While not straying from this fundamental point, this report asks whether it is possible to produce a set of broad pointers that could – as a very rough first approximation – equip non-analysts with a starting sense as to the likely quality of quantitative analytical work.

Are there simple 'rules of thumb' that could orient decision makers when thinking about the likely reliability of analysis? Given the asymmetry between benefit and harm (the latter being easier to achieve and more deeply felt), are there ways of spotting poor analysis that don't rely upon technical knowledge? How might decision makers know when a more detailed, and specialist, review of technical work is needed? Can their intuitions be primed?

To explore these questions, the Strategy Unit interviewed 11 senior analysts (see Annex A). They were selected to provide a breadth of perspectives, spanning: think tanks (health and non-health related), universities, health charities, and health and care services.

In navigating what follows:

- Readers wanting results are referred straight to Section 3. This contains the 'rules of thumb' for decision makers.
- Readers interested in the perspectives of the analysts interviewed (which were used to produce the 'rules') will find these views in Section 2.

# 2. Interview findings: what do experienced analysts look for?

This section presents findings from the interviews. It provides the reader with (some of) the perspectives of experienced analysts: what do they look for when judging the quality of quantitative analytical work? Results are presented thematically, listing tests described by interviewees. With caution, given that the method was entirely qualitative, results are also presented with some sense of the weight of opinion (e.g. 'most interviewees thought…').

## 2.1 Source: *where has this work come from?*

The first assessment of quality raised by interviewees was about the organisation - and individuals - that had produced the analysis. Assessments here were about the interrelated questions of credibility, incentives and trustworthiness.

### 2.1.1 Who has produced the work? Are they credible? What are their incentives?

Every interviewee suggested that they treated the organisation and people who produced the work as a marker of likely quality. Even those that did so tentatively - perhaps suggesting that they didn't want to rely upon 'arguments from authority' – were clear that this was a significant, and quick, way of making early judgements on the work.

As a first order test, several interviewees therefore noted that any report should be clear about who has produced it and for what purpose. This was typically elaborated in terms of the reputation and stake held by the organisation issuing the work:

> *"…knowing that the other think-tanks will do quality assurance and curate the analysis before it goes out on their website. You know that they're trying to upkeep their reputation and their analytical ability. It gives you that extra layer of confidence."*

Another interviewee, giving an example of a recent and high-profile NHS evaluation described assessing quality by examining the track record of the organisation that produced it (in this case, the result suggested a lack of likely credibility).

Most interviewees discussed the likely incentives facing the organisation producing the analysis. The general suggestion here was that – equipped with some prior sense of these incentives – they would calibrate their scepticism in approaching the work. Interviewees making this point typically argued that they were more sceptical if they could see commercial incentives. For example:

> *"…they [a colleague who used to work at a management consultancy] were trying to work out the economic benefits of the [name of public body]. They produced some*

*initial numbers and presented it to a senior person who said, 'This needs a bit more client focus.' They went away and applied more client focus and suddenly the number had doubled."*

A small number of interviewees also noted that charities with a lobbying function also had incentives that should induce scepticism in the reader:

*"I'm not very tolerant of reading public health-style analyses that come up with a really big number for how much the economy will save if…[gives specific disease example]. It fills with me a huge amount of doubt if it's funded by a charity whose main interest is raising funds for their cause."*

### 2.1.2 Has the work come from a peer-reviewed journal?

Around half of interviewees mentioned publication in a journal as a marker of likely quality. The starting sense was that the peer review process could, in principle, guarantee quality. But, in the main, interviewees did not see this as a clear indicator, noting that not all journals are equal:

*"There are now thousands and thousands of journals. Just the fact that something has managed to get into a journal isn't really a mark of any sort of validity."*

Others were more sanguine, noting that credibility could still be assessed, but that this required knowledge of the quality of different journals in a field. For example:

*"There is a massive difference in journal quality. One paper might be published in [gives journal name] another might be published in [gives journal name]. One of those would be the pinnacle of a career…the other I get spam emails from, trying to get me to pay £200 to put my paper in there. But, to a layperson, they sound so similar."*

### 2.1.3 Has the work been 'assured' by a trusted analytical colleague?

The final marker of likely quality relating to its source was analysis coming via a colleague. This was mentioned by around a third of interviewees; in the main, they were pointing to the need for decision makers to have access to trusted advisors. For example:

*"A Chief Executive wouldn't necessarily have, but they should have, a Chief Analytical Officer in their organisation who is curating their reading."*

## 2.2 Transparency: *is the analytical approach clear and explicit?*

The second set of tests raised by interviewees related more closely to the technical content of the work. Within the overall theme of transparency, interviewees raised questions about:

the quality of data and sources; concepts and definitions; use of assumptions; and whether underlying models and data were shared.

### 2.2.1 Is the method well described? Has data quality been assessed? Can sources substantiate claims?

Every interviewee raised the question of methods and data sources. There was little variance in views on this point, and the essence was a desire for clarity and transparency. If this was there, it was treated as a marker of likely quality; if it was lacking, it became a 'red flag':

> *"The methodology element is critical for me…what have you done with this data? How have you processed it? What have you constructed with it? If it's a particular type of analysis, what statistical tests have you done?"*

> *"I just want to know what you did. So, if you're very transparent about the decisions you made, then I can relax. Even if I disagree with some of them, that's actually okay if you've been perfectly clear about what you've done."*

A small number of interviewees noted that they used the presence and length of the methodology section of a report as an 'at a glance' indicator.

Others took more general signals from whether or not the quality of data sources was explicitly assessed:

> *"I'm not even going to get excited about your analysis if you haven't told me what's wrong with the data….if you're not able to talk about the data quality, then that just gets me going: 'Well, I'm a bit concerned about other things - what have you buried or cut corners on?'"*

And, at a level of detail further in, others made judgements based on their knowledge of specific data sources. For example:

> *"There are some datasets where I know, 'That's a really good quality dataset. It's well filled out. It's going to have lots of detail in and it's great to use.' Other areas where I know there's not good quality data…it starts to ring alarm bells."*

In addition to tests of data quality, several interviewees said that they looked at references and sources. The basic question here was whether claims could be substantiated by sources. For example:

> *"There's something about flipping to the back page and seeing where the sources came from. If there's nothing there, that's a red flag."*

*"One of the first things I'll do is check the bibliography because, perhaps a little bit irritatingly, I want to know that they're referencing what I consider to be the canon."*

One interviewee described the process of tracing the footnotes from a report by a global think tank claiming that AI would 'revolutionise healthcare'; the sources used were mainly newspaper opinion pieces and industry case studies.

### 2.2.2 Have concepts been defined? Does the data support the concepts?

Around a third of interviewees mentioned taking conceptual definitions as a marker of quality. For example:

*"If you do a paper on inequality or equity or fairness, say what that means."*

Interviewees making these points also typically noted checking that the data used to examine the concepts in the analysis are at the right level of abstraction and detail. One interviewee gave examples of analysis of health inequalities that used overly broad groupings, such as 'Asian', which masked very relevant differences in outcomes.

### 2.2.3 Are any assumptions explicit and clear? Are they plausible?

Nearly all interviewees mentioned assumptions as a vital element in assessing the transparency of analysis. Several made this point by noting the pivotal effect that assumptions can have on results. For example:

*"What are the key assumptions this rests on, and what judgments are they making that go into their modelling? There was a report a few months ago about the impact of school exclusions where they came up with 'School exclusions have a huge cost to the economy. If we stopped excluding pupils it would pay for itself'. The underlying assumption the authors had made was that if kids didn't get excluded, would go on to earn the average amount. That's one hell of an assumption."*

Having raised the topic, interviewees described what they looked for as markers of the 'honest' use of assumptions. The most commonly raised point was that there should be transparency:

*"Hopefully assumptions will be listed. It's a red flag if they're not."*

Others made more detailed points. A small number of interviewees suggested that having an empirical (rather than a hypothetical) source for assumptions would increase confidence. For example:

*"Knowing that the assumption is rooted in the real world - rather than an analytical assumption which might make sense if you're just doing some numbers, but in the real world doesn't make sense."*

And one interviewee noted that they looked favourably upon analysis formatted such that they could vary assumptions and see how this affected results:

*"I think what [names think tank] do nicely in a lot of their graphs is to have a range. So, you see they've given you: 'What if this assumption was here, or here, or here,' and you get some options. That's nice to know that they've thought about it and shown, if you have put an assumption in, what if you change it slightly?"*

### 2.2.4 Has the underlying analysis been shared? Can it be checked / replicated?

Nearly all interviewees said that they took sharing underlying models / code / data as a marker of likely quality. Support for this approach to 'open analytics' was expressed in different – albeit complementary - ways.

Some noted the problems inherent in not taking an open analytics approach. One interviewee described a recent experience with a company that was brought in to assess potential savings in their area:

*"There was a whole load of analysis and modelling in the background, which we didn't really get access to, and the bits that people did see was full of holes. It was loads of very, very broad assumptions…the data in it has been largely discounted."*

Another interviewee summarised the same point at a more 'in principle' level:

*"There is no way to evaluate trustworthiness if we can't see what you did. Period."*

Others made the case for open analytics by noting the likely effect on people producing analytical work – and therefore the confidence that the reader might have. For example:

*"If someone has put it in the open, they probably have worked to a higher standard, I would assume. But, nonetheless, just putting it in the open doesn't mean it's good. It still requires someone to come along and test it and verify it."*

*"It's definitely a good sign. It's a sign of confidence in the rigour of your analysis and that you're open to challenge."*

## 2.3 Expression: *is the analysis laid out with care?*

The third set of tests raised by interviewees related to the way in which analytical results are set out for the reader. Nearly every interviewee raised points relating to this theme. In

general, points raised were more detailed and specific than for other themes; they are therefore presented more briefly.

### 2.3.1 Has care been taken in the overall presentation?

Around half the interviewees reported using the apparent care taken in the overall presentation of analytical work as a marker of quality. One interviewee noted that they:

> *"Almost feel the weight of it. Does it look like they've invested in the work? Now, that's intuitive to a degree, but in the old days a book would come through, and you'd go: 'Well, they've invested in the publishing, so there's a likelihood that this is going to be important, because it's in hardback."*

Others noted that they used 'look and feel' in a similar way. One interviewee said they judged work negatively if it was set out using the default styling in Microsoft packages, for example.

Another interviewee noted some subtlety within this test. They reported using poor quality presentation as an indicator of poor work; but they cautioned that the reverse was not necessarily true, and that high quality presentation could mask low quality analysis:

> *"Management consultancies often play this game of putting out stuff that looks really slick, PowerPoint presentations, etc…But inevitably, if it's a sloppy presentation, regardless of the content, instinctively, I don't trust it."*

### 2.3.2 Are statistical terms used with care?

Around a third of interviewees noted that they took specific points of statistical expression as indicators of sloppy work. For example:

> *"Are they clear when they're talking about a percentage point change, a percentage rate?"*

> *"If we are quoting percentages, does it contain the numbers as well? Do they have an understanding how many people we're talking about, or is it just fluctuating percentages?"*

### 2.3.3 Has a suitable time period been used? Are data sufficiently current?

Around half of the interviewees made points relating to time periods and the recency of data used. For example:

> *"Is it a recent source? Is it from 20 years ago, or is it from two years ago? Obviously, it depends on what you're looking at. Maybe something from 20 years ago is relevant."*

Several interviewees mentioned the effect of the pandemic on the selection of time periods; for example:

> *"Is it a problem that it's overlapping with the COVID period? Have things massively changed since COVID, so actually, anything pre-COVID isn't very interesting here?"*

### 2.3.4 Is uncertainty recognised?

Around a third of interviewees mentioned looking for recognition of uncertainties in results. The concern here was the way that single-point estimates might affect decision making. For example:

> *"Confidence intervals, not just having central estimates, not just having single values. Because single values invite too much fixation on 'it will be exactly this'. There needs to be an appreciation of the uncertainty."*

This was commonly therefore raised in terms of making forecasts and examining possible scenarios, where the test was about the inherent uncertainties of the future:

> *"When I look at forecasts I always ask the question, 'Why am I not seeing two or three variations of the forecast?' If it's just one line it always surprises me."*

## 2.4 Findings: *how plausible and useful are the results?*

The final set of tests raised by interviewees were about overall findings. Here, interviewees largely described taking a step back and asking about the plausibility of results (are they surprising? Are conclusions supported by the evidence? Are limitations clear?) and their practical applicability.

### 2.4.1 Are results surprising?

Around half of those interviewed reported an increased sense of scrutiny – and scepticism – for results that are on a scale and/or against a trend to a surprising degree. Specific examples given tended to centre around evaluation results. One interviewee suggested that they began from a starting point of disbelief:

> *"The health service has been around for so many years, so if you've got a massive effect size[1], it's not likely to be believable."*

[1] The scale of change following from an intervention, e.g. a 20% reduction in emergency admissions.

They elaborated this point with an example, noting that they would look for a well-described, and plausible, causal pathway connecting a service change with claimed outcomes:

> *"If you're doing an intervention, like integrated neighbourhood teams and you say: 'We took 100 patients and we reduced their readmissions by 30%.' I say: 'These are the same clinicians just working differently…so, do we assume that previously, their care was dreadful? What's happened to these clinicians to change their behaviour?'."*

Two other interviewees made similar points, but from the perspective of results being surprising because they contradicted prior expectations:

> *"It adds a bit of credibility where they also produce results that either aren't that interesting or don't back up their starting point…that tells me that they're not just only reporting the results they want."*

> *"I saw some reporting of an intervention about trying to prevent emergency admissions, and they showed that it backfired. It was reported as: 'Ah, boo, this backfired,' and instead it should have been: 'Look, we proved that this is the wrong thing to do, and now we're going to stop'."*

### 2.4.2 Do findings suggest awareness of the context in which they might apply?

Around half of those interviewed suggested that they assess quality by a test of likely applicability. Here, the test centred upon contextual awareness: does the analysis demonstrate sufficient knowledge of the services / policy that it recommends changes to?

On this point, a small number of interviewees began thinking about the role of non-analysts and decision making (the subject of the next section). Here they suggested that the test of practical application was helpful to improving analytical quality. For example:

> *"The advantage a non-analytical leader has is business knowledge. Sometimes that's really, really important. When they're reading something does it make sense to them? If it doesn't, sometimes you'll find a hole in the report."*

### 2.4.3 Do findings come with a sense of humility and limitation?

Around half the interviewees pointed to trusting more in analysis that demonstrated clear awareness of limitations. As one interviewee said:

> *"Self-criticism is a good indicator that the work can be relied on."*

And another interviewee, making the same point from a different angle, said:

*"I can always tell when [analysis has been] taken over by someone from comms or PR and done a bit more work to it than they really should. That rings alarm bells."*

The prevailing point was that greater trust was placed in results presented with an appropriate sense of humility. This was articulated as being seen both explicitly (e.g. limitations being listed) and implicitly (e.g. a careful, non-polemical, tone). For example:

*"I like people to acknowledge the weaknesses of their approach and the limitations…There's always a tendency for people to over-claim, but I prefer to see a dose of humility."*

### 2.4.4 Do findings go with the grain of prior beliefs? How do findings sit with expert intuition?

Finally, and least tangibly, around a third of interviewees reflected on the role of feeling and expert intuition. The role of feeling was suggested by one interviewee as providing a guardrail against dismissing or accepting findings because they do or don't accord with personal (perhaps political) views:

*"Lots of these things are where someone comes up with something and it's almost too good to be true because they like the findings. It's like the Wigan deal[2]. Everyone wants to believe that…The Tim Harford thing[3], which I love, is when you read the claim to ask how it makes you feel."*

A small number of interviewees described following their intuitive 'feel' for the quality of the analysis. They typically did so by suggesting that these intuitions had been honed over years of analytical experience. For example:

*"There is, it's really hard to describe, but there is a general feel. There's an instinct element to it and I think that absolutely does come with experience."*

[2] See, for example, https://www.kingsfund.org.uk/insight-and-analysis/projects/lessons-wigan-deal
[3] Elaborated here: https://reutersinstitute.politics.ox.ac.uk/news/keep-your-emotions-check-and-three-other-tips-think-rationally-about-data

# 3. ‘Rules of thumb’ for decision makers

This section draws upon interview findings and sets out a series of ‘rules of thumb’ that non-analyst leaders might gain from in making rough judgements about analytical quality. It also provides suggestions for their development and use.

Returning to the analogy with which this report started, one interviewee argued that:

> *“An anaesthetist, how many people could they kill? Three, before someone fires them? Or irreparably harm, maybe, five people? It's countable on one hand. But an analyst who says, 'Stop vaccinating over here, and move over here,' or 'Stop keeping patients with dementia in this ward, and move them out to the community.' You'd have catastrophic health effects on individuals, and then huge impacts on a population level.”*

There is no suggestion in what follows that rough pointers can replace analytical expertise, or established checklists for assessing methodological quality. With the potential for harm (and gain) suggested in the quote above, decision makers clearly need access to high-grade analysts whose advice they can trust.

The ‘rules of thumb’ overleaf are perhaps therefore best thought of as ways of priming the senses, and perhaps therefore:

- Guiding the non-analyst decision maker to a position where they know when to seek more detailed specialist support. The ultimate value of these rules might therefore be in promoting better working relationships between decision makers and analysts; and
- Being better equipped to spot obviously bad examples of analytical work.

What follows should be read in this light.

## 3.1 Rules of thumb

These are set out overleaf as a single page for ease of use.

## As a decision maker being supplied with analytical work, ask yourself:

### *1: What are my starting beliefs?*

- Healthy scepticism is a useful default attitude. And so, before reading the work: how do you feel about the subject of the analysis? Do you want certain things to be true? Be careful if you might be reading to have these prior beliefs confirmed.

### *2: Where does this work come from?*

- Has this come to you via a trusted analytical colleague? Do they think it's good?
- Is the purpose of the analysis stated?
- Is it clear who has funded it? What are their likely interests in doing so?
- Is it clear who has done the work? Are they credible? Do they have a track record in this area? What are their likely incentives?

### *3: Is the analysis transparent?*

- Has the underlying analysis or code been shared? Could it be checked independently?
- Are methods and data sources described in detail? (Do statements of method seem to be hiding anything behind 'innovative' techniques?)
- Can sources be traced and checked? Do sources relate to the context (e.g. not taken from radically different health systems)?
- Are any assumptions clear and plausible? Do they connect with real world experience (i.e. not *'if everyone performed as well as the top 20%'*)?
- Are limitations described? Is there explicit recognition of uncertainty?

### *4: Do the results make sense?*

- Are findings surprising to you? (If so, be sceptical. Especially for claims about *'intervention x causing outcomes y and z'*).
- Do findings fit with your knowledge of the context (e.g. normal patterns of change)? Are findings situated within what is already known about the topic?
- Overall, do you feel as though the analysis is seeking to inform or persuade you? How could the findings inform your decision making?

***Having done this - especially where the analysis is informing a decision - you might want to ask for more detailed review by an analyst.***

## 3.2 Next steps

The rules of thumb above resulted from a short, exploratory piece of work. They are a starting point – intended to stimulate further work - not a final product. To be useful they need to be discussed adapted and adopted. Suggested next steps are therefore to:

**1) Propagate this work**

These initial rules of thumb should be shared on the Health Foundation / Strategy Unit websites, offered to the analytical community as a starting point for discussion: does this seem like a useful way of helping decision makers to approach analytical work?

**2) Invite adaptions and adoptions**

If the basic notion underpinning this project – that it is possible to offer pointers to non-analysts – seems sound, then the rules will be most useful if they are adapted to specific contexts. They can be taken and changed by relevant organisations and networks, such as the Association of Professional Healthcare Analysts; the NHS Chief Data and Analytical Officer Network; the Health and Care Analytics conference; etc.

The Health Foundation and Strategy Unit should therefore invite others to develop, improve and use the 'rules' presented in this report: how can they be made most useful to promoting the use of high-quality analysis?

# Annex A: Interviewees

The Strategy Unit extends its thanks to the following people for giving their time to be interviewed for this project:

1. Ben Zaranko, *Institute for Fiscal Studies*
2. Charles Tallack, *Health Foundation*
3. Danielle Jefferies, *The King's Fund*
4. Jake Abbas, *NHS Humber and North Yorkshire Integrated Care Board*
5. Jess Butler, *NHS Grampian & University of Aberdeen*
6. Kate Cheema, *UCLPartners*
7. Marc Farr, *Kent and Medway Integrated Care Board & East Kent Hospitals NHS Foundation Trust*
8. Richard Wood, *NHS Bristol, North Somerset and South Gloucestershire Integrated Care Board & University of Bath*
9. Richard Wilson, *Birmingham and Solihull Integrated Care Board*
10. Rony Arafin, *British Heart Foundation*
11. Sally Gainsbury, *Nuffield Trust*

Interviews were semi-structured. Interviewees were provided with a briefing note outlining the intention and focus of the work (Annex B); during the interview they were then invited to describe how they assess the quality of analytical work – and whether any of these 'tests' could possibly be performed by someone without technical, analytical knowledge. Interviews were recorded, transcribed and analysed to pick out such tests. Direct quotes, edited for clarity, are used in the report to allow interviewees to speak for themselves as far as possible.

# Annex B: Briefing note for interviews

Leaders need a way of knowing whether analysis merits consideration. On a very practical level: they need to know whether a report is likely to be worth reading in detail, and potentially taking seriously, or not.

In some cases, this will be more or less clear. If they are familiar with the analyst's previous work; or if it has been through peer review or similar, for example.

Frequently it will be unclear. Maybe the work has come from an organisation they are only vaguely familiar with; maybe it comes with the badge of an institution they generally regard as credible; maybe it comes from a consultancy with a recognised name; maybe it is well presented and looks credible, but they don't know the organisation at all; etc.

In these circumstances, how is a decision maker - who is unlikely to be an analyst by background - supposed judge the likely veracity of the work? How would they know whether they can safely discount a report? When they should send it to an analyst for a more detailed review? When they can probably just trust what they have?

This project will investigate how people experienced in reviewing quantitative outputs approach this. It will also examine whether it is possible to produce a small number of tests – simple 'rules of thumb' - that a non-specialist could apply, which would provide rough but reasonable approximations as to the likely quality of quantitative analytical work.

To examine this, we are undertaking a short series of interviews with people engaged in the production / communication of analytical work. We will explore questions such as:

What rules of thumb do you use when approaching unfamiliar analytical work? How do you begin assessing likely quality? What signs do you look for? At the extremes, how do you spot work that is obviously credible or not credible? How do you know whether more detailed analytical attention is needed?

And – crucially - could any of these tests be performed by non-analysts? What rules of thumb could non-analysts (especially leaders) use when it comes to spotting poor quality analytical work?

Interviews will be semi-structured (taking this note as a starting point); be conducted over the phone / Teams; and last around 45 minutes. Any views / tests shared will be treated in confidence and not attributed to any individuals in any reporting.

The output will be a short report outlining interview findings. If it proves possible, this will also include suggested rules of thumb: short and simple tests, clearly expressed and quickly comprehensible to someone with no technical analytical knowledge.

Results will be shared with participants, and made public on the Health Foundation and/or Strategy Unit websites.

**The Strategy Unit**

Email: strategy.unit@nhs.net
Web: www.strategyunitwm.nhs.uk
Bluesky: @strategyunit.bsky.social

This work is licensed under CC BY-NC 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0