To mislead or not to mislead – why preventing misuse of statistics is more effective than combatting it.

Elise Rohan[1]

[1]*Office for Statistics Regulation, United Kingdom*

Abstract

As our world becomes more abundant with data, statistics are increasingly used to persuade and provoke discussion. The UK Office for Statistics Regulation's (OSR) vision is that statistics should serve the public good. Statistics play an important role in supporting democracy, and a big part of that is encouraging their use in wider debate, but it also involves combating and safeguarding against misleading statistics.

This paper explores the definition of misleadingness in the context of our work as a statistics regulator and how preventative measures can uphold public confidence in statistics, enabling misuse to be more easily called out where it occurs. The paper draws from OSR's experience of investigating the potential misuse of statistics and sets out the different ways misleadingness can present in the production and use of statistics, for example, the challenge with complex statistics being distilled into headlines for press releases and social media posts. The paper shares our experience of investigating concerns in the lead up to the UK Government 2019 General Election and how this will underpin our regulatory response to the 2024 General Election.

The repetition of incorrect or unsupported statistics creates a validity through reuse, known as the 'illusory truth effect' or repetition bias. Research on this phenomenon has found that we have a cognitive bias to perceive confidence and fluency as characteristics of truthfulness. Our paper discusses why focusing on reprimanding those that misuse statistics does not drive the system-wide changes needed to prevent it happening in the first place.

We share our findings from our 'intelligent transparency' campaign, which is formed around three principles: equality of access, enhancing understanding, and independent decision making and leadership. These principles are designed to support statistics producers to consider the potential for misuse throughout the life cycle of statistics. We also highlight the power of collaboration and partnerships to call out repeated misuse by individuals.

Finally, the paper takes stock of our learning to date on championing the effective communication of statistics to support society's key information needs. We share our findings on the common pitfalls of communicating statistics that open statistics up to misuse. This includes the presentation of uncertainty and the use of infographics.

Keywords: misleadingness, misuse, transparency, communication, election

Introduction

The Office for Statistics Regulation (OSR) is the independent regulatory arm of the UK Statistics Authority. In line with the Statistics and Registration Service Act (2007), its principal roles are to set the statutory Code of Practice for Statistics; assess compliance with the Code; and report any concerns on the quality, good practice, and comprehensiveness of UK official statistics.

The OSR is responsible for managing the UK Statistics Authority's casework function, undertaking monitoring to identify issues, and investigating issues raised with the Authority. It uses its voice to stand up for statistics and to represent the public, reporting publicly where it has concerns and highlighting good practice. The approach to casework and intervention is guided by a publicly available interventions policy.

The potential for statistics to mislead does not only occur from the use of incorrect statistics. Increasingly, the misuse of statistics occurs from a lack of transparency. Transparency and clarity support public confidence in statistics and the organisations that produce them and minimise the risk of misinterpretation of statistics and data.

Taking a transparent approach by default is particularly important during a general election campaign, where statistics play a role in informing decisions made by the electorate. The OSR does not act to inhibit or police debate, and it recognises that election campaigns require careful judgement about when to intervene. This paper summarises the OSR's work on understanding misleadingness in the context of statistics regulation and how this has informed its approach to responding to the next UK General Election.

1.   Determining the need for a definition of misleadingness

1.1 The role of the statistics regulator during an election

The 2019 General Election was not the first time OSR intervened during an election. During the 2016 European Union Referendum campaign, the Chair of the UK Statistics Authority wrote publicly on the use of the £350 million figure by the Vote Leave campaign. Prior to that, the OSR intervened publicly during the 2015 UK General Election campaign regarding statements on health spending and living standards.

On 28 October 2019, the UK Parliament agreed to hold a snap general election. The OSR's guiding principle for casework during the 2019 UK General Election was to intervene if we thought the statement was materially wrong to the point it could impact on a reasonable person's voting intentions. In the three weeks leading up to the election, the OSR intervened publicly on statements made by politicians and political parties regarding violent crime, homelessness, youth unemployment, and school funding.

The OSR team was proactive in monitoring public debate and online communications during this time. The focus of this monitoring was on issues which the OSR had previously commented on. The monitoring gave regulators a good idea of what would be covered and proactively start highlighting contentious issues early.

The OSR's experience of responding to concerns during elections highlighted that statistics are increasingly used to persuade and provoke discussion. In many cases, statistics are seen as a tool to strengthen weak arguments, and this creates a challenge for the statistics regulator to uphold its role in promoting and safeguarding statistics in public debate.

1.2 Developing a definition of misleadingness

When cases about the use of statistics are raised with the OSR, complainants often use the word 'misleading' to describe their concerns. Judging whether something is misleading is inherently subjective, and it can be difficult to know whether incidents where misleading or incorrect figures are used are deliberate or not. However, the OSR considered that it would be useful to have a definition of misleadingness to guide its interventions and maintain some consistency in its judgments.

In May 2020, the OSR initiated a project with Professor Jennifer Saul from the University of Sheffield to produce a think piece on what misleadingness means in the context of statistics regulation. The think piece presented a series of approaches, explored three methods for tackling misleadingness, and discussed how each one has its own unique benefits. These were:

- Materiality and intention – an approach which focuses on the significance of the statement being made. What were the intentions of the speaker?

- Audience – an approach which focuses on audience understanding. Were the audience misled about what the statistics were telling them?

- Case-based – an approach which focuses on particular features of the presentation of statistics. Is the style of presentation unclear and likely to mislead?

The work resulted in a definition of misleadingness for the OSR as 'We are concerned when, on a question of significant public interest, the way statistics are used is likely to leave a reasonable person believing something which the full statistical evidence would not support.'

As this work took the OSR beyond its usual area of expertise, the think piece was accompanied by a call for feedback to help shape the definition. A strong sentiment from the feedback received was the need to distinguish the production and use of data and statistics. It also found that the intention of the person making a statement was not a helpful basis for guiding or supporting judgements about misleadingness, as determining whether someone has intended to mislead is difficult, subjective, and likely to lead to unnecessary controversy.

This work led to the publication of a follow-up think piece in May 2021, which downplayed intention, recognised complexity, and was clearer on the circumstances in which it is relevant

to consider misleadingness. The think piece redefined misleadingness in the OSR's context as 'We are concerned when, on a question of significant public interest, the way statistics are used is likely to leave audiences believing something which the relevant statistical evidence would not support.'

1.3 Launching the intelligent transparency campaign

The coronavirus pandemic saw statistics and data take an increasingly prominent role in government communications and public discourse. There were many good examples of governments across the UK publishing increasingly large volumes of near real-time data in accessible ways to inform the public about rapidly moving situations. The most prominent examples were the data dashboards published by the devolved administrations and the UK Health Security Agency, and the daily televised Downing Street briefings.

While the OSR recognised the need for governments to make increased use of management information to inform the public, this information was not always equally available to the public, which risked damaging public confidence in the decisions being taken based on the data. The OSR intervened about the lack publicly available data quoted in statements regarding Levelling Up, fuel stocks, hospital admissions, and travel lists. Under the vision that statistics serve the public good, the OSR believes that governments should make it easy for people to understand and scrutinise the data it uses to make decisions and inform the public. This led to the development of the Intelligent Transparency campaign which launched in February 2022.

Intelligent transparency involves proactively taking an open, clear, and accessible approach to the release and use of data and statistics. Intelligent transparency should be the default approach to communicating all statistics, data, and wider analysis. The guidance underpinning the campaign sets out the principles and practices that governments and organisations should follow to ensure the transparent release and use of data and statistics. It is informed by the Code of Practice for Statistics and supports the Office for Statistics Regulation's interventions policy. In practice, intelligent transparency means:

- Data used by government in the public domain should be made available to all in an accessible and timely way.

- Sources for figures should be cited and appropriate explanation of context, including strengths and limitations, communicated clearly alongside figures.

- Decisions about the publication of statistics and data, such as content and timing, should be independent of political influence and policy processes.

Everyone in government has an important role to play in achieving intelligent transparency, not just analysts and statisticians. Departments and organisations need strong analytical leadership, within and beyond analytical professions. For example, communications professionals will support intelligent transparency by seeking advice from analytical colleagues when using data and statistics in public communications. Senior leaders can champion intelligent transparency and create a culture which supports it when promoting the work of governments or reporting on operational activity.

2. Future-proofing our approach

2.1 Reviewing our misleadingness definition

Since the launch of the campaign, the OSR has continued to promote intelligent transparency across government, working with a range of organisations and professions to embed the approach. The principles of intelligent transparency appear to be well embedded in the statistical system, with regular feedback from statistics producers on how the guidance has supported them to challenge decisions around the release and use of statistics by government ministers. However, since 2021, the OSR has seen more cases where the use of statistics is misleading due to a lack of context rather than the figure not being in the public domain. Examples of cases that lack context include those where key caveats are missing, comparisons are made between data that are not comparable, and definitions are not sufficiently clear which affects interpretation of the statement.

In 2023, we commissioned a PhD philosophy student, Kyle Adams from the University of Waterloo, under the academic supervision of Professor Jennifer Saul, to review our definition of misleadingness. The project first reviewed the inclusion of 'significant public interest' in the definition and highlighted that if statistics are used in a way that was likely to leave a small audience with a wrong impression, it would still be misleading, despite not meeting the criterion of significant public interest.

The project then explored the inclusion of the phrase 'likely to leave audiences believing something' and concluded that an audience's belief was not something to be constrained by and that the OSR should take into account the wider body of statistical evidence to decide if a claim being made is misleading or not. The project also explored what is meant by 'audience' and whether what may be misleading to some, may not be to others. For example, when a physicist says that a vacuum sucks in the surrounding air, an audience of other physicists would understand this to mean that surrounding fluid pressure pushes air into the vacuum. However, an audience untrained in physics is much less likely to understand the different

mechanics at play. This analogy led to a conversation about what constitutes an audience and if the focus for the OSR should be something more representative of the public.

The view or belief of the public is important to the OSR. It agreed that its definition of misleadingness should focus on the relationship between the statement being made and the broader body of statistical evidence but maintain a consideration of the view of a 'reasonable person' to articulate judgements in an accessible way. The work resulted in an updated definition of 'We are concerned when, on a question of significant public interest, statistics are used to communicate a descriptive statement that the wider relevant statistical evidence would not support, despite otherwise being an accurate statement.'

2.2 Testing our thinking

As well as commissioning a project to review the definition, the OSR participated and facilitated several roundtables to test its thinking and seek alternative perspectives on misleadingness. The first event was in July 2023 when the OSR participated in a 'Trust in Evidence' roundtable organised by Sense about Science. The discussion considered whether action taken to address misinformation is more harmful than the misinformation itself. One contributor argued that 'fake news' is driven by a lack of faith in democracy which then drives a lack of faith in government-held data.

The discussion also covered 'conspiracy theorists' and considered that these are not always uneducated people. People tend to anchor to things they hear that align with their beliefs, and therefore people often look for a binary true or false response to confirm their view. However, evidence is rarely binary, and this can lead to people distrusting the answer and perceiving the producer as 'not knowing what the right figure is'. To mitigate this, the OSR's approach is to explain rather than rebut in its interventions. This approach allows the OSR to subtly guide people through the data to help them reach its view on the accuracy of the claim, without a direct criticism of the claim or person making the claim.

The OSR's approach was endorsed by a group of philosophers who took part in a roundtable with the OSR in January 2024. The discussion concluded that OSR should focus on being a 'visible referee' by promoting good information – creating a positive conversation around the counter narrative. The roundtable also discussed how the structure of the internet and social media means statistics can gain a life of their own, even when all the right steps are taken by the producer. Social media flattens information to the same level of importance, which means individuals see posts from media outlets and public figures alongside posts from family and friends. Subtly misleading claims can more easily go undetected when individuals are passively consuming information in a series of posts covering a range of topics.

The flat nature of social media creates a particular challenge for government bodies who are already speaking outside of their context when on social media – communicating policy in an abbreviated way to the public. The audience of 'the public' comprises various different groups, and there is a risk that tailoring the communication to the lowest level of understanding opens more opportunities for the communication to be misinterpreted.

These discussions were brought to light in an election disinformation experiment run by the organisation Thinks Insight, which the OSR contributed to. The research found it was more effective to promote good information early rather than reacting to bad information as engagement is the key driver of amplification on social media platforms. Where users comment on a post to question its validity, they are in fact amplifying that content to their networks, which draws more attention to it. The experiment, therefore, found that flagging content as misinformation is not effective. However, inoculation against misinformation, which was tested through a 45-second misinformation detection game, had a statistically significant result on reducing amplification of misinformation.

2.3 Identifying best practice around communication

The conversation around social media is not only relevant to the OSR's work on misleadingness but its approach to casework more generally. Increasingly, the claims it investigates originate from social media and online communications. Social media platforms thrive on simplified and engaging content, which means those speaking on behalf of government often choose to use visuals with short captions to communicate their messages publicly. While this can be advantageous for reaching wider audiences, brevity can often be a source of confusion as key definitions and caveats can be lost in abbreviated claims.

When reviewing concerns about the use of statistics on social media, the OSR guides organisations to consider how easily the social media post can be verified and test how a statement alone might be interpreted if taken out of context. Where information can be shared as part of a thread, this can be a useful mechanism for providing the necessary context and signposting to further information on the statement. Where the risks of a statement being misunderstood is high, the OSR advises that organisations should not attempt to distil the message into a social media post and focus on using the social media post to signpost to another place where the necessary explanation can be included alongside the data, for example, pointing to a press release or a publication.

Regarding data visualisation, the OSR has seen an increase in the use of infographics to present statistics. Infographics can be a powerful way of communicating data quickly and clearly to tell a story. They can, however, create potential to mislead where complex data are

oversimplified, or the appeal of the visualisation is prioritised over statistical accuracy. For example, many of the infographics that the OSR has investigated have contained breaks in axes or truncated axes, which do not start at zero, and risk making changes in data look more dramatic than they are. The OSR advises that axes on charts in infographics should be clearly labelled and any images or icons used to represent different quantities or scales for comparison should be proportionate and the icons relative in size to the changes the image is attempting to describe.

The OSR is continually working to identify opportunities to prevent misinterpretation rather than relying on intervening when things go wrong. In 2022, the OSR commissioned a review of statistical literacy. The review concluded that it is important that statistical literacy is not viewed as a deficit that needs to be fixed, but instead as something that is varied and dependant on the context of the statistics and factors that are important in that context. The OSR recommends that producers of statistics focus on how best to communicate statistics through different means that can be understood by audiences with varying levels of understanding.

Communicating the uncertainty associated with the data is an important part of tailoring communication to different levels of understanding. Different users may want different information about uncertainty depending on the nature of the decisions they're faced with making and their level of expertise. The OSR's report on 'Approaches to presenting uncertainty in the statistical system' published in 2022 found a mixed picture in the UK statistical system.

There are many cases where uncertainty is presented in some form in statistical bulletins – in the narrative, charts, and infographics. Good examples include using words like estimate within the narrative, the inclusion of error bounds in charts, and clear lists of ways that the statistics can and can't be used. There are occasions though where estimates are presented as though they are absolute facts. Not acknowledging that uncertainty could exist within them could lead users to false conclusions. Phrases like 'care needs to be taken' and 'caution is needed' are widely used, but producers could be more specifically helpful in guiding appropriate use of the statistics. The OSR is continuing to develop its regulatory work to better understand what effective communication of uncertainty looks like, and to encourage good practice across government data outputs.

3.   Conclusion

The OSR has been able to turn around casework quickly to achieve impact through the noise of an election campaign in previous years. However, it anticipates that the next UK General Election will be more challenging due to the increasingly online nature of election campaigns.

Based on its work around misleadingness and the lessons learned from the 2019 UK General Election, the OSR is taking steps to improve its channels for intervention to ensure it can react at pace to misleading statements, particularly those shared on social media.

To prevent misleading use of statistics in the first place, the OSR is doing more proactive preparation for the next general election to help people navigate through the various claims and figures quoted in public debate. It is publishing a series of explainers that will cover the common mistakes in public statements that it has seen through its casework across topics which are likely to feature in an election campaign.

The OSR is also working in partnership with other organisations and regulators whose vision is aligned with OSR's and who support the good use of evidence in public debate. The aim is that the collective group contributes to the effective functioning of the election campaign.

References

The Office for Statistics Regulation (2022), The Code of Practice for Statistics Edition 2.1, https://code.statisticsauthority.gov.uk/

The Office for Statistics Regulation (2021), Our Interventions Policy, https://osr.statisticsauthority.gov.uk/policies/our-interventions-policy/

The UK Statistics Authority (2017), Letter from Sir David Norgrove to the Foreign Secretary regarding the UK's financial contributions to the European Union, https://uksa.statisticsauthority.gov.uk/correspondence/use-of-350million-per-week-figure-to-describe-uks-financial-contributions-to-the-eu/

The Office for Statistics Regulation (2020), Misleadingness: A short thinkpiece, https://osr.statisticsauthority.gov.uk/publication/misleadingness-a-short-thinkpiece/

The Office for Statistics Regulation (2021), Misleadingness: A follow-up thinkpiece, https://osr.statisticsauthority.gov.uk/publication/misleadingness-a-follow-up-thinkpiece/

The UK Health Security Agency (2020), UKHSA data dashboard, https://ukhsa-dashboard.data.gov.uk/

The Office for Statistics Regulation (2022), Intelligent Transparency, https://osr.statisticsauthority.gov.uk/transparency/

Thinks Insight and Strategy (2024), Addressing election mis and disinformation in 2024, https://www.thinksinsight.com/reports/addressing-election-mis-and-disinformation-in-2024

The Office for Statistics Regulation (2023), Statistical Literacy: Research, https://osr.statisticsauthority.gov.uk/publication/statistical-literacy-research/

The Office for Statistics Regulation (2022), Approaches to presenting uncertainty in the statistical system, https://osr.statisticsauthority.gov.uk/publication/approaches-to-presenting-uncertainty-in-the-statistical-system/