

Commentary on the [Sturgis Review](#)

Prepared by Better Statistics CIC ([BSC](#))

1. Summary:

This commentary arises from consideration of the review recently conducted by Professor Patrick Sturgis (PS) as commissioned by the Office for Statistics Regulation (OSR). Better Statistics welcome that review and are grateful to the OSR for having commissioned it; we also welcome the subsequent decisions to strengthen the OSR's ad hoc reviews in the future.

We also share the widespread opinion that the ONS showed considerable initiative in setting up the Covid-19 Infection Survey (CIS) so quickly in response to the immediate threats to the NHS, created by the Covid-19 pandemic in the spring of 2020. We nevertheless consider that the subsequent developments of the survey were less assured and that there were failures to re-evaluate many aspects during the subsequent 12 months or so. Although constructive, the Sturgis review does not provide any new evidence to alter this opinion and Better Statistics remain dissatisfied with the lack of information on many of the technical aspects of the CIS, moreover we believe that its description by the Evening Standard and others as a *'gold standard'* study, strengthens the need for full disclosure of all the facts relating to the statistical procedures employed in the sampling and estimation process. The protocol documents are exhaustive in their descriptions of the testing procedures used to establish individual evidence of the existence of infection and we have no concerns with that innovative element of the study.

Accordingly, we describe below those details we believe have been omitted from the ONS and OSR's documentation and also were not covered within Professor Sturgis's review. Although they vary from items that will be of significance to many observers to more parochial concerns, in each case BSC consider that they are worthy of further consideration. We therefore request that the OSR should ensure that the ONS will provide a complete description of the methods used for the survey, addressing the following concerns in detail:

- The original design for the sample covering the reasoning for the definition of the 133 special areas, the actual process used for their construction and the demographic characteristics of each area.
- The *'design-based approach to weighting'*, showing the details of the response rates the weighting is designed to adjust for and other characteristics considered for the weighing. To include specific examples of the weighting and the effects on the infection rates of changes from un-weighted to weighted data. Details of any modifications to the system applied over time and the reasons for such modification, these should include indications of any changes in response rates by the 133 areas and how they were corrected for.
- Consideration of alternative methods - to include descriptions of any trial work; for example testing of alternative invitation letters, use of outbound telephone contact to improve recruitment and any other trials. We are also concerned to understand what efforts were made by the ONS to relate the data arising from the CIS with other data, in particular administrative data available from various sources within the NHS.

- Finally, we consider that it is in the public interest for the ONS to disclose the outcomes of the *continuous improvement* and *benchmarking activities* specified under clause 7 of the contract between the ONS and IQVIA. Hitherto, the ONS has refused to provide such information on the grounds of commercial confidentiality.

Whilst Professor Sturgis's review has confirmed that the ONS should have been better at providing information to the public and that the OSR could improve their review process; his review does not seem to recognise how important it is that the public should expect better of the ONS and, in particular, the OSR. The CIS was not ever spoken of as just another 'ad hoc' survey, it was a flagship study supposedly admired across the world and no-one, other than UKSA insiders and a few cognoscenti, would be aware that it never had the status of providing any **National Statistics**.

When, in August of 2021, BSC had identified an apparent discrepancy in the recently published QMI we had expected to receive a short explanation with, possibly, a correction and that would be the end of it. Unfortunately, that was not the case and the subsequent correspondence between BSC and the ONS shows a continuous failure on the part of the ONS to properly engage with our concerns preferring to obfuscate and to conceal important facts, claiming that to be in the public interest.

That correspondence had been [provided to the OSR](#) by BSC and yet the OSR's review of August 2022 upheld the CIS as fully adhering to the code of practice for all three pillars - '*trustworthy*', of proper '*quality*' and providing '*value*'. It was BSC's concerns with that outcome that had led to the OSR commissioning the Sturgis Review. The OSR is rightly concerned of the risk that it is not perceived as independent from the rest of the UKSA, in particular from the producers of statistics. In the opinion of Better Statistics it needs to be much more robust with its reviews if it wishes to ensure that the perception of independence is to be maintained. However in our view, the current code of practice lacks specific directions on questions of quality and value, sufficient to ensure that the OSR can always perform its duties to ensure the public good.

It appears that the OSR has recognised this concern and in correspondence to BSC they have proposed a review of the code in 2023. We therefore look forward to contributing to that review in due course.

2. Introduction:

Professor Sturgis (PS) confirms that it was the [correspondence](#) between Better Statistics CIC (BSC) and the ONS and OSR that had led to the OSR commissioning his report. In doing so he had highlighted the fact that BSC had raised the following concerns with the ONS and the OSR:

- the methodological approach used for the CIS,
- the quality of the information published by the ONS about the methods used
- the value for money of the CIS
- the adequacy of the regulatory oversight of the CIS provided by the OSR

Better Statistics welcome Professor Sturgis's review as a useful contribution to the question of governance of our statistics, nevertheless we continue to take a different view on many of PS's conclusions as described below.

3. The Methodological Approach used for the CIS:

It should be noted that BSC have never specifically claimed that the methods applied for the CIS have provided erroneous results, rather we have claimed that the information provided by the ONS has been inadequate to make a judgement on the quality of the project and we have observed that the consistent failure to provide information when requested, is cause for concern in its own right.

Although Professor Sturgis upholds BSC's concerns about the failure to provide adequate information, he is confident that the methods applied were probably the best available at the time. His report does not, however, contain any new evidence to support that conclusion. . PS therefore confirms the opinion of the OSR that nothing material could or should have been done to reduce the cost of the project or, possibly, to increase its value.

BSC continue to question this conclusion and wish to highlight the following areas where we believe evidence should be substituted for opinion:

- **The original design for the sample.** On page 2 of the Sturgis report PS says "*The AddressBase Sample was drawn in a single phase, stratified by 133 CIS areas (a bespoke geography created by the ONS)...". BSC had been unable to trace further information on the 133 areas but the OSR have now pointed us to the [web-site mapping](#) of the areas used for the single stage design. Nevertheless this mapping does not seem to provide the information that would justify the effort and cost of creating such a bespoke stratification. We continue to expect an explanation of the reasons for not using a straightforward probability proportional to size sample of administrative areas, stratified by region. As far as we are aware No reason has ever been provided for this special design. Nor have the methods used to create this design have ever been described.*
- **The 'design-based approach to weighting'.** On page 3 Professor Sturgis refers to using a design-based approach to the weighting using parameters considered in the original survey design and then (if we understand correctly) he clarifies the choice of the Bayesian model for the weighting, because of the subsequent integration of the same Bayesian method to estimate infection rates for various subsets of the population including local areas. The primary reporting of this work was provided in the [Lancet article](#) published in January 2021. No further information has been published as far as we know. In our opinion the details underlying these decisions should be published by the ONS to evidence that this innovative trade off has not sacrificed the accuracy of the estimates for the ease of operation of the overall procedure.
- **Other methodological issues:** Some other issues relating to the methodology used by the ONS arise from the comments below centred upon other aspects of PS's report.

4. Alternative methods:

BSC had never specifically proposed any alternative methods, although our correspondence had included questions as to whether there had been consideration of alternative approaches. The ONS had not responded to these questions but the OSR requested that Professor Sturgis should consider whether an alternative approach to the CIS methodology could have been more cost effective. PS's conclusion is unqualified that the methods used by the ONS were the best that could have been expected, given the circumstances. Although we agree that the initial survey work was well considered (with the possible exception of the design of the AddressBase sample into 133 specially defined areas) we nevertheless believe that alternative methods should have been considered after the survey had been running for a year. Of course, it may be that such work was undertaken but that

the ONS consider it to be confidential – as Ms. Emma Rourke explained in her letter to BSC of 20th December 2022 “*the ONS is unable to release further detail because both benchmarking and continuous improvement would require the release of commercially sensitive information*”. Certainly there is no evidence of any changes having been considered until the introduction of the ‘*digital only*’ approach in July 2022.

The main claims made by PS that we question are as follows:

- **Sample Recruitment:** whilst we recognise the good sense evidenced by recalling previous respondents to ONS surveys to get the CIS up and running and the use of the post to introduce the project to potential respondents, we are not convinced that it was necessary to ask them to telephone in to agree to the new survey. At that stage of the work it was not necessary to transfer any personal information to IQVIA, just the name and address. We believe that the information commissioner would have accepted the legitimate interest of sending interviewers to the potential addresses, unless they had specifically requested that they should not be contacted. It can be argued that this would have been a waste of the interviewers’ time without a prior appointment and we accept this argument, particularly given the high level of response to this phase of the survey. However, we question the wisdom of continuing with the same process for the larger, new AddressBase sample. In our opinion the low numbers of individuals actually calling the IQVIA telephone centre should have led to reconsidering the approach, including consideration of a clustered sample in place of a single stage sample. Did the lower response rate for this phase encourage consideration of changes to the methods?
- **Sample design:** BSC had asked the CIS team whether they had considered use of a two stage clustered sample rather than a single stage. Professor Sturgis confirms that the reason for a single stage design was because of the clustered nature of the pattern of infection. However, in our view, the ONS lost a significant opportunity to use the unique features of the AddressBase system to select second stage clusters of, say, 9 addresses according to a specific geographic pattern around a central point. This would have produced a replicated sample design enabling straightforward estimates of sample variance and provided better information on the spread of the disease locally. The clusters would be such as to ensure ease of travel by car within the cluster, the fieldworkers had to use a car for their work, because of the requirement for collecting blood samples. Although it is difficult to assess the overall cost benefit of such a design, it is likely to have been more cost efficient in terms of cost per response and provided more information on the spread of the disease. In passing we note that PS claims one advantage of the within household sample selection as the ability to measure the spread of infection within household. Although BSC concur with that observation, we are not conscious of having seen any analysis discussing that issue. Finally, is it not possible that such clustering might have reduced the numbers required for the ‘*very large fieldworker panel*’ required for the survey? Possibly also, some use of ‘cold calling’ on selected addresses could have improved the response rate achieved, without markedly affecting the costs. Finally, we consider that the original documents inviting “*The Resident*” to contact IQVIA should have been more informative of the manner in which the address had been selected. Although relatively few potential participants are likely to have understood the principles of random sampling, it is possible that many would have been more impressed by the **selection** of their household for the study, instead of using the phrase “*households like yours*”. Specifically, the study design assumed participation of **that** household – not one supposedly *similar* to theirs as emphasised in the invitation.

- **Response rate:** We question what evidence is available to support the assertion that the 13% of responses achieved from the AddressBase households selected can be considered to be *“broadly in line with or even above what might be expected a priori”*? We would argue that it is equally likely that the importance of the work, combined with the substantial financial incentive and the benefit to respondents of having their health monitored, could easily have led to a higher than average response rate and certainly not one significantly lower than average. It seems probable that the unusual recruitment method, requesting that people phone to register for an appointment, was less satisfactory than we would have wished. We would also claim that the failure to actively advertise such an important survey had significantly reduced the potential for a higher level of co-operation. We believe that the CIS (and other important surveys such as the BICS) could have benefitted from more widespread awareness explaining to the public the value of inclusive participation to such surveys, particularly in the early phases. We believe that more than one example invitation letter had been used in the recruitment and that there had been a trial of outbound telephone recruitment we would value publication of the effectiveness of these trials.
- **Bias:** There can be no dispute that an overall response of just 13% of households and approximately 10% for eligible individuals does call into question the possibility of bias in the resultant information. Whilst we recognise that weighting to adjust for differential response rates by demographics should remove any evident sample bias, it is always good practice to investigate non-response, wherever possible. As far as we are aware there was never any attempt to conduct an evaluation of the non-response to the CIS. Further, we do not understand Professor Sturgis’s use of the phrase *“Non-response bias is a property of estimates not of samples and arises when the propensity to respond to a survey is correlated with the survey variable of interest”*. Whilst we acknowledge that the response rate is not necessarily an indication of bias, the lower the response the greater opportunity there is for such bias. Moreover, the paper that PS refers to confirms that the degree of interest in the topic under research often has a bearing on the nature of the response and can result in biased results in some circumstances. We consider it logical to assume that persons with an interest in health matters, particularly personal health, are more likely than others to have telephoned to express an interest in participating in the survey. We are unaware of any attempt to investigate such an effect, if any. Finally, PS’s comments concerning the paper published in Public Opinion Quarterly in May 2017 does not seem to us to be relevant, an average difference of only 1.6% across a number of estimates cannot be assumed to be indicative of the difference that might have been discovered with the infection rate from a higher response rate. Particularly if it had used a less self-selecting process than phoning in to register.
- **Weighting the data - adjusting for bias:** More importantly, we consider that the failure to adjust the AddressBase sample for the bias observed in the sample by size of household, could have affected the accuracy of the estimates. Professor Sturgis notes the concerns BSC had expressed on this issue and acknowledges that *“the lack of a clear explanation about all this on the ONS website and the rather piecemeal way this information was communicated to BS in the ONS correspondence would appear to be the main cause of the persistent failure to close this issue down.”* PS does, however, clarify the reason behind dropping the use of household size as a weighting variable at the introduction of the AddressBase sample, quoting three reasons: *1. it had a negligible impact on estimates 2. there was a desire to make the weighting approach consistent between the design- and model-based estimates (and it could not be used in the latter) and 3. there were concerns about the measurement*

quality of household size, both in terms of the population totals available and the measure of household size in the CIS. This confirms the suspicion that we had from our examination of the original QMI and the data collection documents used for the recruitment namely that IQVIA had failed to ensure that numbers of persons in the household had been correctly enumerated and checked. Nor has there ever been a subsequent attempt to obtain that data, despite BSC having questioned the issue.

- **The effect of lockdown:** BSC have also suggested that lockdown would have masked the significance of household size on the incidence of the Covid infection because office workers and children were confined to their homes. Therefore, we believe that the observation that it had a negligible effect on the original estimates is not surprising. Nevertheless, an effect was noted in the original data, and we understand that it was included as a weighting variable for the initial estimates of infection. We therefore conclude that it was not that 'negligible'.

In short, Professor Sturgis sees "*no cogent reason*" to question any of the methodology used and, of course, the OSR had reached exactly the same conclusion in their reviews of the methodology. For our part we believe that these issues should have been investigated in 2021, particularly when we had drawn attention to them. Of course, without any investigation there could not be a cogent reason, only suspicion.

5. Alternative Survey Designs:

In their correspondence Better Statistics had suggested that the CIS team should have considered a 2-stage sample design, probably using a smaller sample, but one designed around one or other administrative geographies' used by the NHS (probably hospital trust catchment areas). We would then have expected to use some form of ratio estimation to obtain local incidence estimates from the admissions to hospital due to Covid.

It should be emphasised that we never worked on such a design, nor have we been asked about the idea in any depth. In particular Professor Sturgis had not requested we should provide any further detail, he dismisses the possibility for the following reasons:

- **Variation of infection:** PS correctly states that there was no reason to assume that the incidence of the disease in the population would be a constant multiplier of the numbers requiring hospitalisation as a result of the disease. Similarly, there was no reason to assume that it is a constant from one area to another. We had not, therefore assumed that the survey would definitely stop after a while. Rather we considered that we could use a smaller sample of trusts to investigate the variation in the incidence, increasing that number if we are seeing substantial variation across trusts. Our main cost saving would have been achieved by reducing the numbers of samples taken and checked, as was subsequently eventually done.
- **Self-selection:** We do not understand PS's reference to self-selection as being inherent in any proposal to use a design based upon a ratio estimator. The fact is that the CIS procedure itself used self-selection through the invitation to phone in, albeit self-selection from a large random sample rather than the whole population. For the record BSC, would always prefer a genuine random sample selection process for any survey work and the choice of sample method would not be constrained by the estimator used.
- **Coverage:** PS correctly questions that a smaller study relying upon ratio estimation for its estimation of the incidence of disease in the population might not yield the details by demographics as obtained from the CIS. That need not be the case however and, as ever,

determining sample size is a function of the objectives. However, BSC would have experimented with the use of telephone interviews with trained enumerators to collect information from a larger sample, with a selected subsample providing the actual tests, as appropriate.

In summary we would have sought a more agile approach, seeking always to relate survey research data with administrative data at a local level.

4) Did the OSR appropriately review the CIS in its 2022 review?

We are pleased that Professor Sturgis has confirmed many of the observations that BSC had made concerning the failure of the CIS team to provide adequate information on the methods used for the study and the OSR's reviews of the study.

PS nevertheless does not conclude that the ONS had failed to uphold the code of practice, even though they had failed to respond to OSR's requests for improvements. Instead, PS appears to excuse the ONS's failures by suggesting that the OSR requests were at '*too high a level*'; an observation that seems perverse to us. We had understood that the OSR were directing their comments to senior people within the ONS – not undergraduates. However, the quote below from the OSR's publicised response to the Sturgis review, implies that the OSR had always considered the CIS to only require one of their "*smaller and more ad hoc reviews*" and this may therefore explain the Professor's comments and many of BSC's primary concerns with the approach adopted by the OSR. Meanwhile Professor Sturgis has recommended a number of improvements that should ensure that some of the concerns expressed by BSC will not recur in future.

- **How has the [OSR responded](#) to PS's recommendations?:** We quote the introduction from the OSR's response in full "*We accept all the recommendations from the review, noting that for our core assessment and systemic review projects we already largely work in a way that is consistent with Professor Sturgis' suggestions. Our priority will therefore be to consider how to extend these important principles to our smaller and more ad-hoc reviews. We will do this in parallel with considering what further improvements we can make to our existing assessment and systemic review processes.*" Although BSC welcome these intended improvements, we remain concerned that the OSR's second review of the CIS had not concluded, or even suggested, that the ONS had compromised their attention to the code of practice by failing to follow the OSR's guidelines and continuing to withholding relevant information.

5) Value for money:

Despite the careful analysis provided by Professor Sturgis, we remain of the opinion that the OSR's second review of the CIS should have notified the ONS of concerns in respect of the code of practice. In particular BSC believe that the failure to provide more details of the methodology of the CIS had compromised acceptance that the conduct of the project satisfied the '*trustworthy*' pillar. We also consider that the lack of information on the methods creates legitimate concerns with the *quality* of the CIS.

In our previous correspondence with the OSR, Better Statistics had suggested that the code of practice should be strengthened by introducing some specific guidelines and standards to be addressed by the ONS in their QML reports.

Although the OSR had reacted positively to most of the suggestions (including commissioning the Sturgis review), Ed Humpherson had specifically refuted the suggestion that anything further was required to strengthen the 'value' pillar. In his review Professor Sturgis has emphatically endorsed that opinion, saying *"Regarding value for money, there is indeed scope for these matters to be addressed by OSR within the existing Code under: T4.4: Good business practices should be maintained in the use of resources. V1.6: Periodically review whether to continue, discontinue or adapt the statistics. V5: Efficiency and proportionality. These provisions would allow OSR to question the value for money of specific aspects of the survey design and fieldwork, such as the level of incentives paid to respondents, or the use of in-person interviewing rather than respondent self-completion. Whether the CIS as a whole represents value for money as a means of addressing policy needs does not seem an appropriate matter for the OSR to determine."*

Whatever the purposes may be of the various clauses from the code as quoted above, the evidence is that **none** of them were ever considered by the OSR in any of their reviews of the CIS, despite the cost of the programme of well over £1 billion. Nor is there any direct evidence of attempts by the ONS to reduce the cost of the program prior to July 2022, when the survey was changed to a 'digital only', respondent self-administration procedure.

However, Emma Rourke (Interim Director General for Health, Population and Methods at the ONS) had implied that there was considerable work done on issues of the cost and value for money, but that relevant information on the subject has been withheld from the public. BSC had questioned what actions had been taken in respect of *continuous improvement* and *benchmarking*, each a requirement specified in the contract between the ONS and IQVIA and Ms Rourke had responded on 20/12/22 as follows:

"In relation to both requests, the ONS is unable to release further detail because both benchmarking and continuous improvement would require the release of commercially sensitive information. We recognise arguments in favour of transparency and accountability regarding the suppliers contracted with public authorities, and we have demonstrated our commitment to this interest by releasing the bulk of the contract we have with IQVIA, and also by actively publishing our payments to suppliers over £25,000 . This adheres to the inherent public interest in the spending of public money."

BSC find this response to be unsatisfactory, suggesting once again that our questions are irrelevant. Ms. Rourke had previously suggested that important details of the work done were the intellectual property of IQVIA, this is directly contradicted by the contract. Nevertheless pertinent questions remain unanswered.

Finally PS comments that the switch to respondent self-administration for second and follow up interviews in July 2022 *"evinces a concern by ONS to ensure the cost-effectiveness of data collection without adversely affecting data quality"*. Ms. Rourke also quotes this fact as evidence of the attention paid by the ONS and the IQVIA to the issues of 'Value'.

Is it unreasonable to ask why this switch was not made earlier?