

Review of the Office for Statistics Regulation's approach to quality assessing the Covid Infection Survey

Background and Context

In January 2023, the Office for Statistics Regulation (OSR) commissioned Professor Patrick Sturgis of the Department of Methodology at the London School of Economics to undertake a review of its regulatory work assessing the Office for National Statistics (ONS) Covid Infection Survey (CIS). The review was commissioned in response to an on-going correspondence between the ONS, OSR, and representatives of the Community Interest Campaign group, Better Statistics. In this correspondence, Better Statistics (BS) raised a number of critical points relating to the methodological approach, published information about, and value for money of the CIS. They also expressed concerns over the adequacy of the regulatory oversight of the CIS provided by the OSR. The objectives of this review are, therefore, to:

- i. Examine the methodological approach of the CIS, review alternative approaches that could have been taken and assess their suitability compared to the approach that was taken.
- ii. Understand whether OSR appropriately assessed the CIS in its 2022 review and whether this review remains fit for purpose.
- iii. Identify whether there are improvements the OSR could make in their approach to future regulatory work of this nature.

Design and methodology of the CIS

I begin with a brief summary of the design and methodology of the CIS, which is described in much greater detail on the ONS and Oxford University web pages. The CIS was established as a means of monitoring the level of infection from Covid-19 in the non-institutional population of the United Kingdom, including at national (England, Scotland, Wales, Northern Ireland) and regional levels, and by demographic sub-groups. At the time of its introduction, there was a great deal of uncertainty about how many people were, or had previously been, infected by the virus, particularly where infection was asymptomatic. The study was therefore intended to provide crucial evidence on a range of matters relating to infection in order to inform

policies around, *inter alia*, the timing, extent, and geography of lockdown, school opening, national and international travel, inequalities in social and economic impact, and the likely future trajectory of the pandemic. The CIS has also been used to detect new variants of the virus, to assess the efficacy of vaccines, and to inform understanding of long covid.

A key objective of the CIS is to provide accurate estimates of the parameters of interest and, to that end, the sample design chosen was based on random selection of households from a sampling frame with very high coverage of the target population. In the first phase of the survey, from April 2020, the sample design was based on recontacting individuals in households which had previously provided interviews to existing ONS surveys and which had given consent to be recontacted in the future (these surveys also drew their samples from the Postcode Address File (PAF) or AddressBase). Of the 20,276 such households invited, 51% ultimately agreed to take part in the CIS, yielding 22,729 eligible individuals.

Because this phase of the CIS was drawn from samples of respondents who had previously completed ONS surveys, a good deal was already known about them, including demographic information such as age, sex, ethnicity, household tenure, and household size. It should be noted, however, that this information is held only for respondents to the original ONS surveys, which themselves have substantial levels of unit nonresponse.

The stock of respondents to previous ONS surveys is relatively small and was soon expended, so the sample design switched to direct sampling from AddressBase in August 2020. AddressBase is considered the gold standard sampling frame for address-based sampling in the UK, as it has equal levels of coverage to PAF but lower rates of ineligible addresses. The AddressBase sample was drawn in a single stage, stratified by 133 CIS areas (a bespoke geography created by ONS), Local Authority, Postcode, and Unique Property Reference Number. Sampled addresses were mailed an invitation to sign up to the survey by calling a central telephone number and, through that, booking an in-person visit from a fieldworker. Of the 1,400,783 households invited at this phase of the survey, 177,923 registered to take part, a response rate of 13%.

Within participating households at the initial visit, information was obtained by the fieldworker from an adult household member on the number of eligible household members, including those available/willing and unavailable/unwilling to take part in the survey. All eligible individuals aged 2 or above who were available/willing to take part were asked to complete a short questionnaire (completed by parents/guardians for young children) covering symptoms, contacts, and demographic information. Respondents were also asked to provide nose and throat swabs and (for a random sub-sample of those aged 16 or above) a blood sample. These visits were then repeated every week for the first month after the initial visit and then every month thereafter. Respondents were provided with vouchers to the value £50 for enrolment and £25 for each wave of the survey they completed, reducing to £20 from April 2022.

From July 2022, the study also changed from fieldworker home visits to self-completion of swabs, blood tests, and questionnaires (the latter completed online or by telephone). The sample size was also reduced by 25% for swab tests and by 20% for blood tests at this point. Swabs and blood samples were sent to accredited laboratories for testing and the test results were then linked to the questionnaire data.

Estimates of infection from the survey were initially produced using a design-based approach with post-stratification weights adjusting for age, sex, and region. Other variables were also used for post-stratification in this developmental phase, including household size but were later dropped for reasons I will consider later. Estimates of infection were also produced using dynamic multilevel regression and poststratification in collaboration with a team of academics from the universities of Oxford and Manchester. This use a Bayesian modelling approach that produces national and sub-national predictions of infection, with the model predictions post-stratified by age, sex, and region. For a short period, both design and model-based estimates were produced in parallel but from May 2020 the design-based estimates were no longer published to avoid confusion arising from small differences between them.

Assessment of the CIS

The CIS sample was initially drawn from respondents to existing ONS surveys, a strategy which enabled the implementation of a probability design at pace in the extraordinary period at the

start of the pandemic. No more suitable approach to sampling was available at the time. The shift to the gold standard sampling frame for address-based samples, AddressBase, also seems entirely appropriate when the initial sample source was exhausted, given the lack of suitable alternatives. A sample of individuals might possibly have been drawn from the list of NHS registrations but this is known to have poor coverage for some demographic groups, such as younger males, and is an individual rather than a household level frame, so would not have provided information on within household infection.

One might question the decision to draw the sample in a single stage, when a clustered design might produce efficiency savings due to the reduced costs of interviewer visits to sampled households. However, as noted in Peter Benton letter to BS of 10/6/22, the decision to use a single stage sample was taken due to the highly geographically clustered nature of covid-19 infections. While the statistical rationale for this decision is not provided in the letter or elsewhere, it seems prima facie appropriate and I assume it was carefully considered by ONS statisticians. My own correspondence with ONS indicates that the very large fieldworker panel deployed for the CIS, allied with stratification by the 133 CIS areas, meant that clustering the fieldwork would not have resulted in significant cost savings while serving to increase variance for some estimates.

Trained fieldworkers were used to list out the household roster, conduct interviews, take blood, and oversee respondents performing swab tests. Again, this represents a gold standard approach and it is hard to conceive of a strategy for this stage of the fieldwork that could be expected to produce higher quality data. The switch in July 2022 to respondent self-administration after the first interviewer visit to households, including a parallel run period where both approaches were used and tested against each other, evinces a concern by ONS to ensure the cost-effectiveness of data collection without adversely affecting data quality.

The most serious threat to the accuracy of estimates from the CIS is the high rate of nonresponse and the possibility of potentially large biases that this introduces. Indeed, this has been the primary methodological concern of BS throughout their correspondence with ONS/OSR. The response rate during the first phase of the CIS is reported as 51% at the household level on the ONS website, although this is rather misleading because it does not

account for nonresponse to the initial surveys from which respondents to the CIS were drawn. Assuming an average response rate of approximately 50% to these initial surveys gives a net household response rate of 25% for the first phase of the CIS, which is considerably closer to the second phase household response rate of 13%. It is worth noting that the individual level response rates published on the ONS website are also, no doubt unintentionally, rather misleadingly overstated because they are given as the number of responding individuals over the number of eligible individuals in responding households. An individual level response rate would usually incorporate an estimate of the number of individuals in nonresponding households, which here would imply a response rate closer to 10% rather than the 91% (for England) reported by ONS.

While these household and individual level response rates are low in absolute terms, and it is therefore appropriate to be concerned about the potential for bias, they are not in my assessment lower than should be expected given the nature of the survey and the context in which recruitment was carried out. It is well known that response rates have been in steep decline since at least the 1980s and prior to the pandemic, interviewer administered surveys in the UK struggled to break 50%, even for relatively straightforward surveys on topics of general interest. The limited number of surveys that have returned to face-to-face fieldwork since the onset of the pandemic indicate that response rates have been adversely affected by the experience of the pandemic, in ways that are not currently well understood, with the standard expectation for response rates now somewhere around 30-40% depending on the design and topic of the survey. Given, the burdensome and invasive nature of the CIS (requiring self-swabbing and blood tests for many respondents) alongside recruitment taking place during a highly infectious viral pandemic, a 13% response rate is broadly in line with, or even above, what might be expected a priori. This is not to say that representativeness is not a concern but, rather, that it should not be considered a result of poor design or implementation by ONS and its partners.

One feature of the CIS design that might have yielded a higher response rate relates to the mode of first contact with sampled households. In the CIS design, this is through a letter requesting that an adult household member contact ONS to arrange an appointment for an interviewer to call at the address. An alternative approach would have been to send an

interviewer to make the first contact, as would be standard in most household surveys and it seems likely that such an approach would have yielded a higher response rate. This was not possible, however, because the data sharing agreement that did not allow ONS to share address information with the data collection agency IQVIA. Therefore, IQVIA could only start the fieldwork when the household registered by directly contacting IQVIA. Even without this constraint, there may also have been considerations related to lockdown restrictions in place at this time that would have made in person visits unfeasible. However, the rationale for this part of the design is not explained in the documentation on the ONS website, nor in the study protocol document on the Oxford University website.

That said, it is my understanding that many of the fieldworkers undertaking visits to households (in the initial stages at least) were not experienced survey interviewers with expertise in contacting households, making appointments, and converting these to interviews. Rather, they were trained in the procedures required to collect the biological data from households and this may have been the reason that this approach to making initial contact was chosen. This may also have limited the feasibility of deploying these fieldworkers as the first point of contact for respondents in the survey timeline. Be that as it may, it seems unlikely that this or other marginal changes in the design of the fieldwork would have resulted in substantial increases in response rate and, still less, the representativeness of the achieved sample. The cost-effectiveness and value for money of interventions that nudge the headline response rate up by a few percentage points are increasingly being questioned by survey methodologists and this is likely a case in point.

It is also important to acknowledge that, even when a survey has a low response rate, it will not *necessarily* produce biased estimates. Nonresponse bias is a property of estimates not samples and arises when the propensity to respond to a survey is correlated with the survey variable of interest (Groves, 2006). This can be seen from the equation below, where the magnitude of the bias in the mean of the survey variable, \bar{y} , in the responding sample is a function of the covariance between the mean of the survey variable and the propensity to respond to the survey, σ_{yp} , divided by the mean of the response propensities of the sample elements, \bar{p} (and where \bar{p} is equal to the response rate for the survey),

$$\text{Bias}(\bar{y}_r) \approx \frac{\sigma_{yp}}{\bar{p}}$$

Holding σ_{yp} constant, the magnitude of nonresponse bias in \bar{y}_r increases as \bar{p} decreases. In general, σ_{yp} is unknown so we can usually only say that the *risk* of nonresponse bias increases as the response rate declines. What this equation also shows is that we should not assume that a low response rate will automatically result in nonresponse bias. Indeed, recent studies have shown that the correlation between response rate and nonresponse bias is much weaker than has hitherto been assumed (Groves & Peytcheva, 2008). For example, Sturgis et al (2017) compared survey estimates across multiple surveys measured after different numbers of calls made to addresses by interviewers across a number of household surveys in the UK. They found an average difference of just 1.6 percentage points between the first call (when the average response rate was just 14%) and the final call (when the average response rate was 63%).

This is not to argue that response rate is not an important indicator of potential bias but rather that it is eminently possible for low response rate surveys to yield approximately unbiased estimates. Another relevant factor to bear in mind here is that the key measure of policy interest from the CIS is *change* in infection over time. In this regard, even if the estimate of the *level* of infection is somewhat biased it seems reasonable to assume that estimates of change would be approximately unbiased, on the basis that there seems no strong reason to expect σ_{yp} to vary much over time.

In addition to random selection, the CIS mitigates selection bias through statistical control and weighting adjustment. In the first phase of the survey, estimates were produced using a design-based estimator with post-stratification weights derived from the joint population distribution of age, sex, household size, and region. This weight also incorporated design and nonresponse weights from the initial ONS surveys and an attrition weight to account for dropout between the initial ONS survey and the CIS. Information about this additional weighting is not reported on the ONS website.

Following the initial phase of the survey, ONS moved to the Bayesian modelling approach, developed in collaboration with world-leading statisticians at the Universities of Oxford and Manchester. This uses covariate adjustment in the multi-level model and post-stratification of the sub-national estimates. The post-stratification variables for the model-based estimates were restricted to age, sex, and region because the joint distribution (which is a requirement for this method) for additional variables was not available at the sub-national level.

ONS also publishes estimates from a separate modelling exercise, the primary objective of which is to understand the characteristics of people testing positive for covid at the national level. The model is unweighted and the predictors included are sex, ethnicity, age, region, urban or rural classification of address, deprivation percentile, household size, and whether the household is multigenerational.

The differences between the procedures used to adjust estimates for nonresponse in the design- and the model-based estimators has been the source of some confusion in the correspondence between BS and ONS/OSR, particularly relating to the (non)use of a measure of household size. From my reading of this correspondence and additional communications with ONS, I have established that household size was used for the design-based estimates in the initial phase of the CIS but was dropped in October 2020 for 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 seems a cogent rationale and I see no strong reason to think that household size should have continued to be included as a weighting variable after October 2020 (I assume that the reason adjusting for household size makes little difference to estimates of infection, even though it is sometimes significant in the unweighted national level model, is because age and household size are strongly correlated and predictive of covid infection). I would add that there are good reasons for keeping a stable set of variables in a weighting matrix, as frequent changes would run the risk of confounding real with methodological change. That said, 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 the issue down.

Overall then, the design, fieldwork implementation, and estimation approach of the CIS are, in my assessment, of a very high standard. No survey is perfect of course and even surveys of the highest quality are prone to a range of random and systematic errors. However, the specification and procedures followed in the CIS seem to me in nearly all respects to be those which are most likely to minimize the mean squared error from all sources. I do not consider that significant cost savings could have been achieved without incurring a negative impact on the volume and quality of the information obtained about patterns and trends in covid-19 infection.

Alternative designs

There are myriad ways in which the basic design of the CIS could be tweaked or amended at the margin, some of which are mentioned above. However, so long as the approach is to rest on random selection of households or individuals and voluntary provision of swabs, blood samples, and questionnaire responses, none could be expected to make a notable positive difference to the overall costs and errors of the survey. This is because the great majority of the cost is determined by factors that vary little, or not at all, across any such design, namely fieldworker pay and expenses, respondent incentives, and test processing. Here, therefore, I am concerned with designs that depart more radically from the CIS approach of random sampling and voluntary provision of data by respondents.

One such radically different approach would be to use the data collected from routine covid testing in hospitals as a baseline from which to estimate the population total of infections. This would have the significant benefit of using covid tests that are taken for another purpose, saving on (what I assume are) the largest fixed costs of this part of the CIS. Indeed, this is a method suggested by BS as a more cost-effective alternative to the CIS in their letter to OSR of 16/3/2022, where they refer to the approach as a ratio estimator. I am not aware of any detailed exposition of the methodology proposed here and am only able to go on the very brief description of it by provided to me by BS in my correspondence with them. My assessment of this method may not, therefore, accord exactly with what BS have in mind.

That said, the basic approach is to use the ratio of the number of patients in hospital and the number of covid infections in the population estimated by the CIS as a basis for projections from hospitals to population infection totals. Clearly, the CIS is necessary to calculate this ratio in the first place but once it has been obtained over some defined period, it is no longer required and the substantial cost of the survey can be saved.

There are, in my opinion, at least two severe limitations of such a method that mean it would not provide a suitable replacement for the CIS. The first is that the ratio of patients in hospital to covid infections in the broader population will likely be subject to possibly quite large fluctuations over time. There does not seem any good theoretical reason to assume this ratio would be time invariant, not least because the composition of people in hospital is subject to quite substantial seasonal variation, as well as being prone to exogenous shocks. It is also possible that changes in the behaviour of the virus, for example new variants, would affect this ratio in ways that are hard to predict in advance. It is, therefore, an approach that might work for a while but then wouldn't and, without the benchmark of the CIS, it would be impossible to know when a change in the ratio had occurred. If the primary methodological concern about the CIS is representativeness, it seems counter-intuitive to switch to a design which relies on untestable assumptions about a highly self-selecting population.

A second major weakness of such an approach is that it would only provide the headline measure of covid infection and would therefore yield little or no data on the demographic breakdown of covid infections, on antibodies, long covid, and the other valuable information that is obtained from the CIS. Nor would it provide information on repeated testing on the same individuals over time. It might be possible to collect some of this information via smaller bespoke surveys but it is difficult to assess the feasibility and cost-effectiveness of such an approach without a detailed description of it and such a consideration is beyond the scope of this report. Moreover, it needs to be remembered that once the initial investment was made in the CIS, any efficiency saving from switching subsequently to a completely different approach would have been much less than the full cost of the CIS. This is not to indulge in the sunk cost fallacy but simply to note that much of the cost of establishing a survey like the CIS is fixed relative to the marginal cost of each interview once it is up and running. In short, I do

not consider that an approach based on projecting national level infections from the number of people in hospital would be a feasible alternative to the CIS.

Another way that the methodology could have been implemented differently from the CIS is through the use of non-probability sampling. If combined with reduced or no monetary incentives for participation, an approach based on self-selecting samples could have met at least some of the information requirements but at a lower cost than the CIS. For example, the [Zoe health study](#) produces regular estimates of population infection in by recruiting self-selecting samples of the public who enter information about symptoms and test results in an app.

However, such an approach has well-known limitations relating to the representativeness of self-selecting samples and the inability to measure asymptomatic infection. And, while this kind of method is able to collect information on attitudes and behaviour, demographic characteristics, symptoms, and long covid, it cannot include blood samples and does not have access to the actual results of covid tests, which require self-report by respondents.

Of course, these parts of the data collection could be added to a self-selecting sample design but this would defray any cost savings as the same processes would be required as for the CIS, irrespective of whether the sample is drawn using probability or non-probability methods. An approach based on self-selecting samples would therefore, in my assessment, produce less data, of lower quality and coverage while delivering only modest cost savings, depending on the exact nature of the design.

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

The OSR first reviewed the CIS in May 2020, shortly after the survey was established at the height of the first wave of the pandemic. This was a light touch ‘rapid’ review that was supportive in tone and highly favourable in its assessment of how the survey had been designed and delivered by ONS in a very compressed and challenging timeframe. Even at this stage, however, OSR drew attention to several areas where it urged ONS to focus its attention on ensuring that the survey meets the standards set out in the Code of Practice for Statistics (henceforth the Code). Notably, this included the need to maintain good response rates and

to be mindful of the different audiences ONS communicates to about the survey, including avoiding overly technical language for non-expert users. There does not appear to have been a published response by ONS to this review.

The second OSR review of the CIS was published in March 2021, with its assessments structured according to the three pillars of the Code: Value, Trustworthiness, and Quality. It again commended ONS for its work on the CIS, particularly its collaborative approach with project partners and the devolved administrations and its contribution to public understanding of the pandemic. The March 2021 review reiterated the need for ONS to be clearer in the communication of its work and future plans to users and raised a number of new points that it asked ONS to address. For the sake of parsimony, I will not list all the recommendations here but instead pick out two for illustrative purposes, as these most pertinently reflect the issues that BS raised with ONS. These recommendations were that ONS should:

- *“improve its published information about methodology and consider how best to communicate this to different types of users”.*
- *“publish information about the representativeness of the survey – for example, what it is doing to increase participation and how the modelling approach accounts for variation in response rates”.*
- *“publish information about the demographics of all participants, to help users understand variation in nonresponse”.*

While these (and the other) recommendations seem appropriate and likely (if implemented) to improve the quality of the CIS and its outputs, I note that they are specified at a rather high level of generality and do not include a timeline for implementation. This makes it difficult for an external observer to assess whether they have been satisfactorily addressed. For example, it is not clear what would or would not satisfy the recommendation to improve the published information about methodology, nor when this action should be achieved by.

A response from ONS to the March 2021 review was not published until May 2022, more than a year later, and it is not clear from the published correspondence on the OSR website, how

this timeline was determined. Notwithstanding the difficult circumstances of the pandemic, it seems a considerably longer time lag than would be desirable for completing and reporting on some of the recommendations. In its response of May 2022, ONS addressed each of the findings and recommendations in the 2021 OSR review, detailing the steps it had taken to implement them and setting out its plans to take further measures in the future. These responses are, in my assessment, relevant and appropriate but ONS is afforded rather a lot of latitude to define what the appropriate actions should be in each case. This is perhaps not surprising given the high level at which the OSR recommendations are made in the review and the ambiguity over what successful completion would constitute.

Parenthetically, I would add that the OSR recorded its findings and recommendations in a summary table, while ONS replied in the form of an open text letter. This makes it more difficult than it needs to be for interested parties to align the ONS actions with the OSR recommendations they are intended to address. There does not appear to have been any published response from OSR to this ONS letter, nor any updating of the publicly available version of the review document to record progress (or lack thereof) following its initial publication.

The third OSR review of the CIS was published on 30th August 2022. It was based on interviews with ONS staff, analysis of survey documentation, and engagement with users. The findings of this review were again very positive and supportive, commending ONS for its strong commitment to the Code and the progress made on developing the CIS statistics since the 2021 review. The motivation for carrying out the 2022 review is stated as being the July 2022 change in the design of the CIS, from fieldworker visits to respondent self-administration of swabs, blood samples, and questionnaires and a reduction in the sample size. This might reasonably be taken to imply that the 2022 review would not have been undertaken in the absence of these changes in the survey design. From this follows the question of how progress against the recommendations of the 2021 review would have been monitored and reported on in the absence of the 2022 review.

The 2022 review is again structured according to the three pillars of the Code, with a mix of positive/supportive and constructively critical findings under each pillar. From my own

reading of the survey documentation, these seem to me to be appropriate assessments of the strengths and weaknesses of the CIS, with critical comments mostly focused on ONS' engagement with users of the survey and the completeness and appropriateness of the information it shares about the survey design and fieldwork. It is notable that the same points raised in the 2021 review regarding the need for more and better information about the methodology of the CIS, the representativeness of the survey and how this is being addressed through statistical adjustment are repeated in the 2022 review, albeit with somewhat different focus and emphasis:

- *“there are some areas for improvement on transparency of development plans and published methods/quality information”.*
- *“methods information of interest to expert users is not always kept up to date”.*
- *“ONS should publish more about the representativeness of the survey. For example, ONS hasn't said why it doesn't adjust for deprivation or similar measure in its methods article”.*

Considering that a year had passed since these points were raised in the 2021 review (indeed the issue of response rates and representativeness was first raised in the 2020 review), it does not seem unreasonable to expect them to have been addressed by this point. So, the OSR review process does not appear to have sufficiently robust processes in place to monitor and enforce progress against its recommendations. From the published record, ONS appears to have been left to define whether, how, and over what time frame it chose to address the recommendations. This apparent weakness in progress monitoring, enforcement, and reporting is exacerbated by the lack of detail and explicit timeline in some of the OSR recommendations noted earlier. It is not clear what the consequences, if any, are for ONS of failure to comply with recommendations in the OSR reviews.

The 2022 review also noted persistent problems with the way ONS makes information about the survey available to users, a point which had also been made in the 2021 review:

“it can be challenging to find information users need – there is a lot of content, and it is not always very joined up (e.g. links to methods information, blogs and external study protocols). To note, we did highlight navigation of the website as a challenge in our last review”.

Having spent a considerable amount of time navigating these disparate documents myself for the purposes of this report, I can fully endorse this characterisation. I would add that a cross-cutting problem with the information provided by ONS is the confusion between the audiences intended for different parts of it. ONS has often been criticised (including by OSR) for providing information that is too complex and technical for non-expert users. However, complex, detailed, and technical information *does* need to be made available in order that expert users can properly understand and assess the methodological procedures and choices that have been taken. The disparate and disjointed information that ONS has provided about the CIS has also mixed and conflated the level of expertise of users it appears to be aimed at, leaving the expert user dissatisfied and the non-expert user confused. In my view, ONS would have benefited from being provided with a greater level of detail in the OSR reviews about the problems with the way it makes information available to users, including clearer directions about what needs to be changed and over what time frame.

At the time of writing, the only published response to the 2022 OSR review from ONS on the OSR website is dated 7 September 2022. This says nothing more than “We are grateful for your recommendations and will use them to inform our continual improvement of these statistics. I look forward to updating you on our future progress against the recommendations you have made.” My experience of producing this report leads me to conclude that little or nothing has been done to address the recommendation relating to information provision since it was made seven months ago and there is no record of progress monitoring or updates on the OSR website.

These matters have an obvious bearing on the correspondence between BS, ONS, and OSR that served as the stimulus for this report. It is clear that the recommendations in the 2021 and 2022 OSR reviews, and the actions taken by ONS to address them were not sufficient to ensure that the requisite information regarding sampling design, fieldwork procedures, response rates, and representativeness was placed in the public domain within a satisfactory

time frame. While I cannot say this with certainty, it seems likely that the difficulties experienced by ONS and OSR in resolving the matters raised by BS could have been brought to a swifter conclusion had OSR been able to ensure that its recommendations were satisfactorily addressed by ONS within a reasonable time frame.

In addition to methodological issues, BS also raised the question of whether the Code should be amended to include specific reference to value for money and methodological guidance for survey practice. On these points, I am persuaded by the rationale for not including such provision set out by Ed Humpherson in his letter to BS of 28/4/2022. 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.

As for including methodological guidance in the Code, this would not be straightforward because best-practice in survey methodology is fluid, often quite context-specific and subject to disagreement amongst practitioners. It would therefore be challenging to include advice of this nature in the Code in a way that is sufficiently comprehensive and up-to-date to be useful. I agree therefore that methodological advice and guidance is best delivered through the Analysis Function Support of the GSS.

In sum, my assessment is that the 2022 OSR review of the CIS (and the two reviews that preceded it) did a good job of identifying the strengths of the survey and of noting where improvements were needed. I was also impressed by the open, collegial, and collaborative

way that the OSR engaged with the queries and criticisms raised by BS. However, there were weaknesses in the ways that the findings and recommendations of the review were communicated to ONS and in how progress against objectives was monitored and enforced. The result of this was that the actions taken by ONS in response to the reviews was insufficient to remedy the issues that had been identified as requiring action.

Improvements in OSR's future regulatory work

It is fair to say that the findings of this report suggest that, in the case of the CIS at least, the regulatory procedures have been insufficient to ensure that full compliance with the Code is achieved within a satisfactory time frame. The problem here lies not so much in the findings and recommendations themselves but in how compliance with them is monitored and held to account. My recommendations are, therefore, that the OSR should:

1. Ensure that the recommendations in its reviews and assessments provide more explicit guidance on the actions that need to be taken by the producer and with a clearly specified time frame for completion for each recommendation.
2. Require that producers publish a response to a review within six weeks of the review's publication, setting out how it intends to act on all recommendations in that review.
3. Monitor and report on progress against the recommendations in its reviews by regularly updating the review Annex, in which the findings and recommendations are enumerated against the pillars of the Code. Review reports should be considered 'live' documents, with progress against recommendations updated when milestones and deadlines become due.
4. Implement a penalty for failure to comply with review recommendations within the specified timeline. This might be a requirement by the head of the unit being reviewed to write a letter to the National Statistician explaining why the recommendation has not been actioned.

REFERENCES

- Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646–675. <https://doi.org/10.1093/poq/nfl033>
- Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(2), 167–189. <https://doi.org/10.1093/poq/nfn011>
- Sturgis, P., Williams, J., Brunton-Smith, I., & Moore, J. (2017). Fieldwork Effort, Response Rate, and the Distribution of Survey Outcomes: A Multilevel Meta-analysis. *Public Opinion Quarterly*, 81(2), 523–542. <https://doi.org/10.1093/poq/nfw055>