Pete Benton | Director General for Health, Population and Methods

Tony Dent
c/o Better Statistics CIC
(*via email only*)

4 February 2022

Dear Mr. Dent,

Thank you for your correspondence with my colleagues Ruth Studley, Deputy Director for the COVID-19 Infection Survey and Esther Sunderland, Head of Strategic Development for the COVID-19 Infection Survey, regarding the methodology of our survey (CIS). I understand that your primary concern is that larger households may be overrepresented in the CIS sample, and that the incidence of COVID-19 infections in the population may be overestimated.

As Ruth and Esther have explained in your previous exchanges, the possibility of including household size into our models was considered in the early phases of the survey, but this information was no longer available once we started using the AddressBase for our sample. Nevertheless, our post-stratification approach on other characteristics and the random sampling ensure that households sampled are at least approximately representative.

Census 2021 outputs will gradually be published over the summer, giving us the opportunity to further improve our approach with the latest available data. Following this, we will consider updating our methodology and I would be happy to inform you of any changes as they occur.

We want to thank you for your interest in our survey and appreciate the feedback you have provided.

Yours sincerely,

Pete Benton

From: Tony Dent <tony@betterstats.net>
Sent: 16 March 2022 17:00
To: Parliamentary Unit <Parliamentary.Unit@ons.gov.uk>
Cc Others
Subject: Re: ONS response to Tony Dent c/o Better Statistics

Dear Ms Hagan,
I have received no reply from Mr. Benton or anyone else to my request of 4th February, as copied below. Meanwhile I have today sent the attached letter to Mr. Ed Humpherson.
I would be obliged if you could pass a copy of today's letter to Sir Ian Diamond with my compliments.
Best wishes,
Tony Dent

From: Tony Dent <tony@betterstats.net>
Sent: 04 February 2022 16:36
To: Parliamentary Unit <Parliamentary.Unit@ons.gov.uk>
Cc: Others
Subject: Re: ONS response to Tony Dent c/o Better Statistics

Thank you, Ms Hagan,
Can you kindly ask Mr. Benton to send me copies of the tables and other information that confirms the assertion in his letter?
With all good wishes,
Tony Dent
Director, Better Statistics CIC

From: Parliamentary Unit <Parliamentary.Unit@ons.gov.uk>
Sent: 04 February 2022 15:45
To: Tony Dent <tony@betterstats.net>
Subject: ONS response to Tony Dent c/o Better Statistics

Dear Mr Dent,

Thank you for your correspondence to the Office for National Statistics, regarding the COVID-19 Infection Survey.

In response, please find attached a letter from Pete Benton, Director General for Health, Population and Methods. I hope that you will find this letter to be insightful.

Kind regards,
Jo-Anna Hagen | Parliamentary and Correspondence Officer | she/her
UK Statistics Authority| Awdurdod Ystadegau'r Deyrnas Unedig
jo-anna.hagen@statistics.gov.uk | www.statisticsauthority.gov.uk

Pete Benton | Deputy National Statistician for
Health, Population and Methods

Tony Dent
Better Statistics CIC
(*via email only*)

01 April 2022

Dear Mr Dent,

Thank you for your follow-up email to my letter of 4 February. I understand that you have raised similar points regarding the COVID-19 Infection Survey (CIS) in separate correspondence directed to my colleagues Owen Brace, Director for Communications and Digital Publishing, Bill South, Acting Deputy Director for Research Services and Data Access, and the Infection Survey Analysis team. Please allow me to apologise for the delayed response.

We appreciate the questions you have raised, which are of importance to many of our users. We will respond to these questions with as much clarity as we can provide and are grateful in advance for your patience whilst we work through any issues we mention. I will address your questions in turn.

1) Household size in CIS sample

As you state in your letter, your main concern is that household size is not adjusted for in the positivity and Incidence models from the CIS.

Since sampling from the AddressBase in early Summer 2020 we do not have population level estimates of household size including by region, age and ethnicity in order for us to include household size in our post stratification for our positivity and incidence models.

Later this year, we will link CIS data to Census 2021 data in which we will have household size distributions for smaller geographies. This will allow us to look into using household size in our post stratification for our positivity and incidence models. We will be transparent about the consequences of this exercise and communicate any changes we make as a result of it.

You asked Mr South to clarify certain terms from the newly launched metadata catalogue[1] via the Secure Research Service (SRS):

- **Dvhsize**: this variable is for those who were sampled from other Office for National Statistics (ONS) surveys. It would therefore have a very high missingness- especially when looking at the most recent data. It is also likely to be out of date now as household sizes change over time and that variable provide a total household size at the time they took part in the other ONS survey so most likely end of 2019 start of 2020.

---

[1] https://ons.metadata.works/domain/index.html

- **N_participants**: this is the count of survey participants and it is not a true household size number.

2) AddressBase sample methods

You requested further details on the method of sampling used when the AddressBase file was introduced as the sample frame for the CIS. I will address these in turn:

- *A description of the key geographical variables used to sort the sample frame:* the sampling frame is sorted in hierarchical order: CIS Area, Local Authority, Postcode and Unique Property Reference Number (UPRN)
- *The details of the CIS Areas:*
    - There are 133 CIS Areas for the UK
    - In England, if a Local Authority population is 200,000 or more, then the Local Authority is the CIS Area. If a Local Authority has a population of less than 200,000, it is merged with adjacent Local Authorities to form a CIS Area
    - In Wales and Scotland, CIS Areas are groupings of the Health Boards
    - In Northern Ireland, CIS Areas match the Health Trusts
- *The statistic used for determining the probability proportional to size calculation – whether total population, total households, total addresses or other:* we use total addresses in each CIS Area to identify the size
- *What, if any, clustering was used within the address sample selection:* we invite multiple people from the same address but do not include any other clustering in our sample selection. This is in line with other ONS Social Survey sample designs.
- *Details of how frequently fresh samples are drawn:* new samples were drawn weekly but no new samples have been drawn since December 2021
- *Details of any variation by CIS Area and the effects of response rates experienced:* we closely monitor our response rates and the characteristics of those responding. We publish information on the representativeness of the survey in our QMI[2]. Over the coming few months we are moving to a digital data collection approach and will be monitoring response rate even closer. Balancing our priorities during this period, we will endeavour to publish more information on this area of the survey methodology when possible.

You also note that the OSR review of the CIS of March 2021[3] stated that *"users told us they would like more information about response rates to fully understand how the characteristics of those who choose to participate in the survey impact the estimates."*

As mentioned, we will publish information on this in the future. However, we are confident our survey is representative due to the sophisticated weighted and poststratification we use, which will also be updated when we have Census 2021 population estimates.

3) Household enrolment

---

[2]https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/coronaviruscovid19infectionsurveyqmi
[3] https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-iain-bell-ons-covid-19-infection-survey-statistics/

You requested copies of the survey protocol documents used to instruct agents conducting in-person enrolment interviews as well as those used when potential CIS participants telephone to express their interest in participating.

As this material is the intellectual property of IQVIA - our partner for the CIS- we are unfortunately unable to share it.

4) Public engagement

Since participation in the survey is by invitation, any blanket advertising to encourage the public to take part could incur significant wastage and confusion with individuals potentially looking for ways to mistakenly volunteer. Our engagement approach to ensure robust response rates is twofold: firstly, to maximise participant retention in the survey via a programme of regular communications and updates and secondly, to reinforce the value of the survey and the participant contribution across all ONS communications channels including digital and social media, media relations and stakeholder engagement.
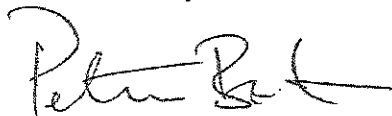
In radio, TV and press interviews, ONS spokespeople consistently thank participants and encourage others who are invited to participate. This includes on major national media outlets such as BBC, Radio 4 and Five Live as well as smaller regional media outlets. For instance, Sarah Walker said on BBC News on 30 March 2021 that *"If you do get a letter from us, please do seriously consider about taking part"*. On 9 July 2021, Sarah Crofts told GB News that *"if any of your viewers are participants please do continue"*.

5) Infection Survey Analysis

You recently requested from our Infection Survey Analysis team the most recent data breaking down the incidence of the disease into symptomatic versus asymptomatic by ethnicity. We currently publish monthly analysis on the proportion of positive cases with symptoms in our characteristics article[4].

I hope this letter has addressed some of your concerns regarding the CIS. I am aware that you have also been in touch with the OSR. They will consider and respond to the concerns you have raised with them in due course.

Yours sincerely,

Pete Benton

---

[4]https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveycharacteristicsofpeopletestingpositiveforcovid19uk/19january2022

Better Statistics CIC

Pete Benton
Deputy National Statistician for Health, Population and Methods
Office for National Statistics

By email via: Parliamentary.Unit@ons.gov.uk

13th April 2022

Dear Mr. Benton,

Thank you for your letter of 1st April. I realise that you will be disappointed by this response to that letter, but we share your opinion that the issues we have raised over the past 6 months are important and we do not feel that your replies satisfy our concerns. Those concerns extend beyond the amount of money that the CIS has cost, to an attempt to ensure that we can all agree as to what best practice should be on future occasions, when a new survey is to be developed. For convenience I use the same headings as your letter and my previous correspondence.

1) Household Size in CIS Sample:

We are puzzled by your statement that "*we do not have population level estimates of household size including by region, age and ethnicity in order for us to include household size in our post positivity and incidence models*".

Although you have not published any details of the weighting algorithms used for the CIS, the use of Bayesian regression methods implies to us that the weighting can be interpolated from the various data items at the disposal of the ONS. These include estimates of numbers of households and numbers of residents by household size and region as provided in the publication
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/datasets/householdsbyhouseholdsizeregionsofenglandandukconstituentcountries.
Similarly, estimates of numbers of households by household size and ethnicity has normally been available from the Labour Force survey.

We would agree that an investigation of our concerns using the Census data is likely to be more accurate, but it will be far less timely. Moreover, the concern that Better Statistics has consistently expressed, namely that the incidence of positivity is likely to be influenced by household size, is confirmed in the January 2022 report that you have now drawn attention to.

We return to that report further below, but we remain dissatisfied by the apparent failure to have investigated the potential for bias introduced by the use of the AddressBase sample, without changing the estimation processes developed from the previous sample methods.

I say *'apparent'* because there has never been any comprehensive description of the weighting and other processes used to determining the CIS figures, a fact that may have led to a misunderstanding on our part.

Meanwhile, thank you for confirming details of the variables **Dvhsize** and **N_participants** in the metadata catalogue. Can you confirm that there is no variable in the file that gives the total number of persons in the household?

2) AddressBase Sample methods:

Thank you for providing more detail on the sample design adopted in the summer of 2020. We assume that more detail will be published at some future occasion; but in the interim we have the following observations:

a) We note that the areas used for Northern Ireland were matched to the NI Health Trusts and for Wales and Scotland they used a combination of local Health Boards. Was any NHS data considered for the sample design in England?

b) Is it possible to have the actual list of the 133 areas used?

c) Why was no clustering used in the survey sample design to make the survey more cost effective? It was known from the beginning that multiple face-to-face calls would be required at selected addresses and clustering would have been a sensible approach to help reduce costs, particularly given the *Value for Money* requirement in the contract with IQVIA.

d) The information provided in the QMI report of July 2021 does not provide any serious detail on the response rates achieved. That report also implies that the weighting for the survey is done at a regional level and not at the level of the CIS areas. We understand, however, that the modelling work is done at the CIS area level, although no details have been published. When do you expect to update the QMI report?

e) If possible, we would like to understand what is meant by moving to a *'digital data collection approach'*? The main subcontractors for the face-to-face work have all used CAPI systems for more than 25 years and the published protocol documents for the CIS all imply the use of computer-based data collection. What is expected to change? Using cloud-based server CAPI interviewing will speed up matters a little but does not, of itself, enable more response rate analysis.

In regard to your confidence that the procedures used ensures that the CIS estimates are representative of the population in the UK; I am only sorry that you are unable to provide the information to support that assertion. To say that you will, in future, provide more information to users about response rates, as requested by the OSR one year ago, does not inspire users with the same confidence you express.

3) Household enrolment

Your assertion that the material we have requested is the intellectual property of IQVIA contrasts with all my experience of contracts for the provision of survey services and I have therefore now reviewed the Contract with the IQVIA for the provision of Covid-19 Infection Survey Services:

Clause 37.1 states: *All Intellectual Property Rights in any material (including specifications, instructions, plans, drawings, patterns, models and designs) which is either:*

a) *Provided or made available to the contractor or the staff by the Authority; or*

b) *Created by the contractor or the staff in relation to the provision of the service and provided to the Authority as a deliverable under the contract*

Is the property of the AUTHORITY. The contractor shall not and procure that the staff shall not use or disclose any such material without prior Approval except where it is in the public domain.

Other clauses in the contract make it clear that the contract conforms with the Authority's obligations under the Freedom of Information Act. Accordingly, I am now hopeful that you will comply with my request, as originally made some months ago, or will I need to make an application under the Act?

4) Public Engagement:

Thank you for advising us of the attempts to engage the public. Personally I still believe that properly advised advertising would assist the ONS to obtain more cost effective responses to the various surveys it undertakes. Meanwhile, your reply gives the impression that you are entirely satisfied with the present position of a 13% response rate to the CIS. I'm sure that is not the case. Or is it?

5) Infection Survey Analysis:

Thank you for advising us of the monthly reports, and in particular the January 2022 report. It was nevertheless difficult to find the information that unpicked estimates of the asymptomatic from the symptomatic prevalence of the disease; although it is evident that the variation in the symptoms that people experience can make it difficult to create a precise picture. Is that the reason why these two proportions are not given greater prominence in the ONS releases?

Included in the January 2022 report is the observation that *"In rural villages those living in larger households were more likely to test positive than those in smaller households; in more urban households, there was no relationship between household size and testing positive."* This provides further evidence that the distribution of the sample by household size should be investigated for potential bias in the results. Meanwhile, you have not responded to any of our requests for data analysis by Household size and none is provided in that report.

Finally, you refer to our recent letter to the OSR. As you may have seen. that letter refers to the possibility of including some more specific reference to *Value for Money* within the Code of Conduct. Value for Money is, of course, included as a requirement in the contract with the IQVIA and we are therefore hoping that you could advise us as to the status of the following aspects of that contract:

a) What specific actions have arisen from the performance monitoring specified in clause 7.2 (c) with the aim of continuous improvement.

b) Also, under Clause 7 of the contract, the Authority is entitled to appoint a Benchmarking Partner. Has the Authority exercised that right? If so, what organisation or individual has been appointed and with what result?

Finally, we propose to publish this correspondence on the Better Statistics website. Please advise if you have any objections to that.

With all good wishes,

Tony Dent
On behalf of Better Statistics CIC.

Better Statistics CIC

Better Statistics CIC

Pete Benton
Deputy National Statistician for Health, Population and Methods
Office for National Statistics

By email via: Parliamentary.Unit@ons.gov.uk

9th May 2022

Dear Mr. Benton,

Further to my letter of 14th April, I have now received a reply from the Director General of the OSR. As a result, I propose to send the OSR copies of my correspondence with the ONS, concerning the Coronavirus Infection Survey.

Before I do so, is it possible that you might clarify some aspects from my letter of the 14th April 2022?

With all good wishes,

Tony Dent
Director, Better Statistics

Pete Benton | Director General for
Health, Population and Methods

Tony Dent
(*via email only*)

10 June 2022

Dear Mr Dent,

Thank you for your response and your ongoing engagement on the COVID-19 Infection Survey. I apologise for the delayed response.

Given that the Office for Statistics Regulation (OSR) will be undertaking a review of the COVID-19 Infection Survey (CIS), I will defer many of your points to the outcome of this review, as I am aware that you are engaging with them on this matter.

However, I will answer a few of your questions directly.

1) Household size in CIS Sample

Dynamic Bayesian multi-level regression post stratification is used for the main positivity modelled estimates from the CIS. Using this technique we adjust for potential non-representativeness over time in terms of sex, age and region.

In order to post stratify by any variable, we need to know the distribution of this variable in the general population and how this varies by other variables in the model. Therefore, we need the number of different household sizes in the UK and how household size varies by age, sex and region in order to include this in the post-stratification.

This is why the household level data from Census 2021 is the only way we can adjust appropriately for household size. We have committed to investigating this when we have the data, later this year.

2. Addressbase sample methods

a) National Health Service (NHS) data was not considered as part of the CIS sample design for England. The geographical variables used to sort the sampling frame for England are the CIS Areas; these are either single local authority areas or a combination of local authority areas depending on the population size of the local authority. There are 133 CIS Areas for the UK.

b) We will endeavour to publish the 133 CIS Areas.

c) Clusters of households weren't selected as the nature of transmission of infectious diseases may cause clusters in the population that are or are not infected. If the design were clustered, there would be a chance that we would be capturing these clusters rather than gaining a fuller understanding of the population.
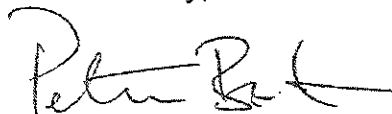
It is worth noting that every individual lives in a household. Since household transmission is a key route of spread, random selection of households should give an unbiased estimate of population level positivity. On the other hand, this may not be the case for sampling clusters of households because of higher very localised rates of transmission. By drawing a systematic sample of households within each strata sorted by various variables related to geography, it gives us an even spread of the sample of participants across each strata. It therefore avoids introducing error through local-level clustering, which may skew results. While value for money is an important factor in designing the survey, this has to be balanced with the need to produce high quality, accurate estimates.

d) We will endeavour to publish more information regarding response rates.

e) A 'digital data collection approach' is in reference to the move from a study-worker led data collection approach to an online data collection approach, where participants complete the questionnaire using a newly developed website. Under this new approach participants will be sent swab and blood kits in the post and be required to return them via Royal Mail or courier.

3. Household Enrolment

This matter has been referred to the relevant team. We will provide you with an update on the matter in due course.

Yours sincerely,

Pete Benton

Better Statistics CIC

Suite 626, Kingsgate
High Street,
Redhill, RH1 1SG
www.betterstats.net

Better Statistics CIC

Pete Benton
Deputy National Statistician for Health, Population and Methods
Office for National Statistics

By email via: Parliamentary.Unit@ons.gov.uk

20th June 2022

Dear Mr. Benton,

Thank you for your letter of 10th June. I had heard that you have been unwell and I hope that your letter signifies that you are now returned to health.

I now consider our correspondence closed, subject only to the matter of household enrolment, but I would be grateful if you would agree that we might publish the correspondence on our website.

Meanwhile, I propose to make the following observations to the Office for Statistics Regulation concerning your reply:

1. Household Size in CIS Sample
   a) I cannot claim to have any experience of Dynamic Bayesian multi-level regression but when we first introduced Bayesian methods into post stratification weighting we were using rim-weighting because our primary (Bayesian) justification was to use all available relevant data, subject only to Occam's razor. With rim weighting, of course, it is not necessary to have the distribution of every variable within all other variables, rather we are assuming that, probably, the sample itself provides the best estimate of such a distribution.
   b) We understand that, when the details become available from the 2021 Census, you will be in a position to improve your estimates but we are disappointed by the failure to examine other possible procedures and the failure to provide more detail of the methods actually employed.

2. AddressBase Sample Methods.
   a) We understand your explanation for not clustering the sample although it would, of course, not have been difficult to have identified the contribution to the variance in the infection rate from within and between clusters. Beyond that a clustered design could have provided other benefits, if carefully exercised, particularly over time. However, our concern was the potential gain in cost efficiency that clustering could have led to, yielding a higher response rate than 13%.

b) Further with regard to clustering, the design employed for the CIS has already allowed some clustering within households, although we are unable to identify any adjustment for this fact within the estimation processes. We would therefore prefer to see more detail on the precise methods applied before we could be convinced that clustering was not preferable. Was it actually considered at the time?

c) In respect of the digital data collection you have mentioned, BSC are increasingly concerned that the methods employed are failing to represent a substantial part of our population; but we will take that issue up with the OSR.

3. Household Enrolment.
   a. We expect to take this up with the OSR unless we hear further from the team.

Thanks again for your reply.

Yours sincerely,

Tony Dent
Director Better Statistics CIC

Cc Office for Statistics Regulation