SCOTTISH HOSPITALS INQUIRY

Bundle of documents for Oral hearings commencing from 19 August 2024 in relation to the Queen Elizabeth University Hospital and the Royal Hospital for Children, Glasgow

Bundle 21 - Volume 7
Substantive Core Participant responses to Supplementary Expert Report of Sid Mookerjee

Table of Contents

1.	A49807974	Greater Glasgow Health Board - Response to Supplementary Expert Report of Sid Mookerjee	Page 3
2.	A49860374	NHS National Services Scotland - Response to Supplementary Expert Report of Sid Mookerjee	Page 16

SCOTTISH HOSPITALS INQUIRY

REVIEW BY NHSGGC

OF

SUPPLEMENTARY REPORT FROM SID MOOKERJEE

DATED 7 AUGUST 2024

1. A report from Mr Sid Mookerjee, dated 9 May 2024, was disclosed to core participants (the "**Initial Report**"). With reference to Scottish Hospitals Inquiry Direction 5, Appendix B at para 2.1, NHSGGC provided specific questions to be asked of the report's author, and specific comments on the substance of the Initial Report, on 19 June 2024. Counsel to the Inquiry provided instructions to Mr Mookerjee to prepare a supplementary report in light of that response. Those instructions were provided by way of a Note dated 12 July 2024 (the "**CTI Note**"). Mr Mookerjee provided a supplementary report dated 7 August 2024 (the "**Supplementary Report**"). NHSGGC was invited, in accordance with Direction 5, to provide comments on the Supplementary Report. This document is NHSGGC's response to the Supplementary Report.

2. NHSGGC's response to the Initial Report noted fundamental and basic errors regarding Mr Mookerjee's analysis and the underlying data used in his comparator exercise. These errors entirely invalidate the findings and conclusions of the Initial Report. The primary issue identified was that the denominators used in the calculation of infection rates were different for RHC than for the comparator hospitals. As the Initial Report did not compare like with like, the conclusions were fundamentally flawed. NHSGGC submits that, notwithstanding the Supplementary Report, these fundamental and basic errors remain.

Summary of response to Supplementary Report

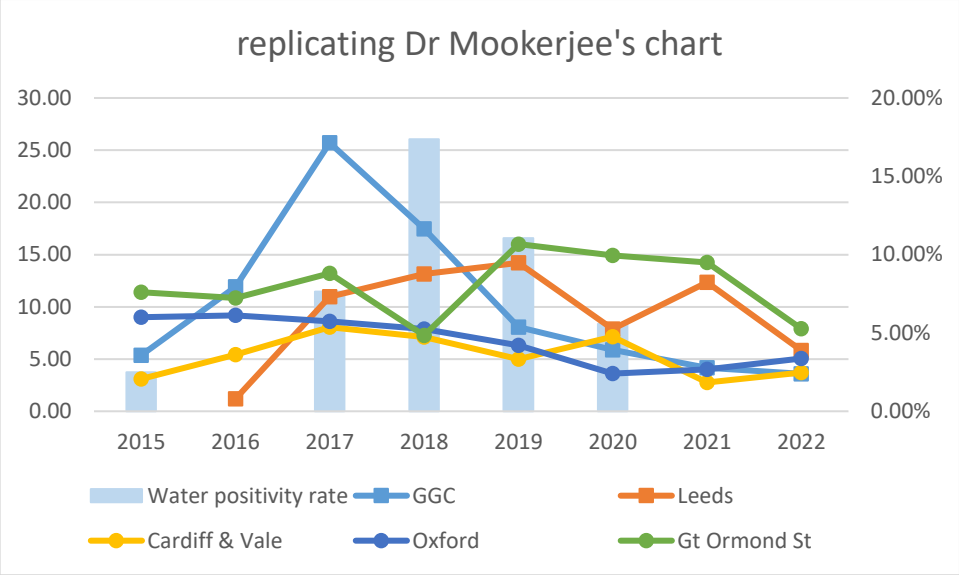
3. NHSGGC submit that Mr Mookerjee has not responded to the comments and questions set out in NHSGGC's response to the Initial Report. There remain two fundamental issues with Mr Mookerjee's reports which result in flawed and unsafe conclusions.

   3.1. The first fundamental issue relates to the data that are used as the denominator for analysis and comparison of infection rates. NHSGGC submitted in its response to the Initial Report that the comparator hospitals had included all day cases and the NHSGGC data included only admissions with an overnight stay.

3.2. In retrospect, NHSGGC acknowledge that day case activity should have been included. However, NHSGGC consider that the Inquiry's data request lacked clarity on the inclusion criteria. In any event, NHSGGC's response to the data request made clear that the data only included admissions with an overnight stay. Whilst the scope of the dataset was specifically stated, it also ought to have been clear to Mr Mookerjee when interpreting the data that they did not include day cases. The magnitude of difference in admissions (compared to the number of cases in each hospital) ought to have made it obvious that Mr Mookerjee was not comparing like with like. The number of admissions included in the NHSGGC first dataset were a tenth of those of GOSH and Leeds comparator sites and a fifth in the case of Oxford and Cardiff.

3.3. NHSGGC have since provided a second set of data that includes day case activity to ensure comparability with the other comparator hospitals used. However, in his Supplementary Report, Mr Mookerjee fails to take most of those data into account. He specifically excludes from his analysis the two wards with the majority of day cases admissions without an overnight stay (Wards 2B and 6A Day Unit). He does so on the purported basis that the new data "doubled, tripled and in some cases increased by 11 times the original figure provided for a given year" (para 2.62 in the Supplementary Report). It is precisely because day cases were included which make up the majority of admissions in these wards.

3.4. By excluding the wards with day cases, Mr Mookerjee's re-analysis was only undertaken using wards 2A and6A (inpatients) which are mainly admissions with an overnight stay. Accordingly, his comparison continues not to compare like with like. The comparator hospitals include day cases, whilst the vast majority of RHC day cases have been excluded. This repeats the error in the Initial Report and the conclusions reached by Mr Mookerjee remain unsafe for the reasons set out in NHSGGC's response to the Initial Report. Mr Mookerjee concludes in his report that the infection rates in the RHC were greater than the comparator hospitals for all the years being analysed. This conclusion is unsafe as a result of the fundamental differences in the dataset being compared. NHSGGC consider that it is vital that this mistake is rectified by Mr Mookerjee by the inclusion of admissions to wards 2B and 6A (Day Unit). NHSGGC are prepared to assist with that process should the Inquiry wish that. NHSGGC have undertaken the analysis the Inquiry requested of Mr Mookerjee, to include all day cases (as the comparator hospitals did). This comparison shows that NHSGGC's infection rates are generally similar to or lower than the other hospitals, for most of the years in question. Our re-working of the graph demonstrates the year on year variation seen across all hospitals and that similar relative changes can be seen in other hospitals.



3.5. The second fundamental issue is that NHSGGC consider that the statistical analyses carried out by Mr Mookerjee are flawed. NHSGGC has reviewed the supplementary report and their comments in this regard are set out below.

4. NHSGGC emphasises that these are fundamental errors and not points of detail. They ought to be clear to anyone with a basic understanding of statistics, and not just to those with the expertise claimed by Mr Mookerjee. The nature of these fundamental errors in the basic statistical analysis calls into question Mr Mookerjee's reliability as an expert witness. NHSGGC's internal experts in epidemiology, statistics and microbial ecology have analysed the report and have provided detailed commentary below. NHSGGC have provided the credentials of its internal experts at Annex 1. NHSGGC are prepared to provide further detailed analysis, should the Inquiry desire it to do so. However, if independent analysis is desired by the Inquiry, NHSGGC urge the Inquiry to instruct an independent statistical review of Mr Mookerjee's analysis. NHSGGC consider that, given the fundamental issues with Mr Mookerjee's analysis, any recommendations made by the Inquiry based on his analysis, will be unsafe.

Detailed analysis of Supplementary Report

5. NHSGGC has noted particular questions from the CTI Note below in bold, followed by commentary on Mr Mookerjee's response to those questions. These responses focus on the fundamental issues summarised above. However, these fundamental errors call into question the entirety of Mr Mookerjee's analysis.

e) Explain the basis of the approach to correlation analysis at paragraph 10.2 and how it is possible to use a small number of data points to draw conclusions regarding a correlation between water positivity and infection rates.

6. Mr Mookerjee's answer (paragraphs 2.28-2.29 of the Supplementary Report) is objectively incorrect. Mr Mookerjee chose to collapse the large data sets provided by NHSGGC by aggregating them by year. This left him with only eight data points for infection rate and six for water positivity rate. Correlation is run on pairs of values so he was left with only six pairs (listed in the table following 10.2 in his original report, p.38), and he further reduced this by excluding the values from 2020, leaving only five pairs of numbers, i.e. five data points. The amount of data used to calculate each of these yearly rates is irrelevant to the correlation analysis itself, its robustness, or the level of confidence in the data. Running a correlation analysis on only five data points is significantly flawed. Basic statistical textbooks recommend 25-30 data points, at a minimum. For a correlation coefficient of 0.7 to reach statistical significance ($p<0.05$), a minimum of 13 data points would be required.

7. Furthermore, presenting the correlation coefficient from an analysis with five data points, without any kind of confidence interval or p-value, and claiming anything about the 'strength' of the correlation, is significantly flawed. With such a small number of data points, there is a high probability that random chance alone has generated the reported correlation coefficient. Where such a non-significant correlation coefficient computed from so few data points is used, any claim of 'moderate to very strong' positive association is flawed and unsafe. This is a fundamental issue with Mr Mookerjee's analysis and calls into question his credentials as an expert in this area. Mr Mookerjee argues that confidence intervals and p-values are not required as these are real-world data. We disagree with this as explained in paragraph 31 below.

i) Explain whether marked increase in water samples taken in 2018 after small number were taken in 2015 and 2016 may be in response to the IMT investigations and what (if any) biases might this introduce considered in the interpretation of the data? Thereafter to consider whether the correlation between water positivity and infection rates may be affected by such bias and was this considered in the interpretation of the correlation analysis?

8. Mr Mookerjee's answer at paragraphs 2.38-2.40 of the Supplementary Report is not relevant to this question. Mr Mookerjee has simply repeated the sections from his initial report that explain issues and challenges with raw data sets. The following paragraph, 2.41, also does not address this question, but repeats Mr Mookerjee's explanation for the exclusion of the 2020 data from the correlation analysis. Mr Mookerjee claims that he excluded this data point because '2020 saw a 20% drop in water testing, compared to 2019', and because of changes

'in the hospital context owing to the onset of the Covid-19 pandemic'. This explanation is insufficient to justify exclusion of the 2020 data. From the table following paragraph 8.4.19 in his initial report, water sample numbers per year are as follows: 80 in 2015, 47 in 2016 (a 41% decrease), 196 in 2017 (a 317% increase), 1158 in 2018 (a 491% increase), 1809 in 2019 (a 56% increase), and 1469 in 2020 (a 18.8% decrease, which Mr Mookerjee rounded up to 20% in his explanation). It is clear that: (i) water sampling numbers remained high in 2020 (the second highest annual number after 2019), indicating that water testing continued throughout the pandemic period despite the severe challenges faced by NHSGGC; and (ii) at 18.8%, the percent change in sample numbers year-to-year is lowest from 2019 to 2020. Far greater changes in sample numbers occurred between all other years, and yet none of the other data points were excluded. Furthermore, the raw data sheets from which Mr Mookerjee computed these values clearly show that the sampling locations within the wards and the types of tests carried out in the laboratory did not change in 2020 despite the pandemic.

9. The exclusion of the 2020 data has two impacts: (i) it artificially and wrongly forces the trend of water positivity upwards, a trend that is already forced artificially upwards by what Mr Mookerjee states are underestimates in 2015 and 2016, and (ii) the correlation coefficient increases from 0.54 (with the 2020 data) to 0.66 (without the 2020 data), though neither correlation is statistically significant. Excluding data from an analysis without suitable justification is a fundamental flaw with Mr Mookerjee's analysis and calls into question his credentials as an expert in this area.

10. Mr Mookerjee addresses the question of bias in paragraphs 2.43 and 2.44, but his explanation is self-contradictory. In 2.43, he posits that, had sample numbers been higher in 2015-2017, the 'water positivity figures would have been more representative and in line with the evidence […] from Dr Walker's and Dr Mumford's papers', and 'we would have seen a higher overall level of water positivity'. But he expresses water positivity as a percentage, so if the true 'water positivity' was indeed so high, then increasing sample numbers should not result in a jump from 0% to 17.4% positive as this would have been detected even with small sample numbers. Despite stating that the values computed for 2015-2017 might not be 'representative' and implying that they would have been higher with larger sample numbers, in the very next paragraph (2.44), Mr Mookerjee states that there is no bias concern in assessing the 'trend in water positivity' over this period. Both positions cannot be true. Either the earlier periods are underestimates, in which case any trend analysis is clearly biased, or the earlier periods are considered representative, and a trend analysis is justified. Mr Mookerjee's explanations show a poor understanding of bias.

k) Re-run the comparison eventually presented in the Quantitative Report at 9.7 and 9.10 using Grand Total column in the First Admissions Data Set as the denominator for the purpose of calculating rate of infections per 1000 admissions.

11. Mr Mookerjee has not answered this question. Simply stating that it would be 'inappropriate' to do so, without a full explanation, is not an acceptable answer.

l) Re-run the comparison eventually presented in the Quantitative Report using 'Occupied Bed Days by Ward' … instead of admissions

12. Mr Mookerjee has not answered this question, instead referring to earlier explanations for why he initially used admissions. This is irrelevant. The Inquiry specifically asked Mr Mookerjee to re-run the comparison with occupied bed days. Mr Mookerjee did not fulfil this request.

m) Use data in the summary table at paragraph 8.3.6 of the Quantitative Report to create separate magnitude charts (in the form of the chart after paragraph 9.7) and comparator BSI rate per 1000 admission and 1000 Occupied Bed Days Per Ward for each of the comparator hospitals (…) so as to discover whether there is any significant difference in the rates of infection… The results of this exercise should be presented in graphical form…

13. Mr Mookerjee has not answered any of the parts of this question. The Inquiry specifically requested that new magnitude charts be computed against each comparator hospital, using both BSI rate per 1,000 admissions and per 1,000 occupied bed days per ward, and that the results be presented in graphical form. Instead, Mr Mookerjee copied his original table 8.3.6 into the Supplementary Report, and copied his original trend graph, largely unchanged (only with trend lines removed and individual comparator lines also shown). There are no new magnitude charts (in the form of the chart after paragraph 9.7) and no new computations based on bed days. These charts still show the NHSGGC BSI rate without day cases versus the comparator hospitals' BSI rate which include day cases. This is plainly not comparing like for like and artificially increases the incidences of infection within the RHC.

n) Use the Second Admission Data Set (including patients admitted with an admission date who did not stay overnight) to repeat his analysis with that new data and draw conclusions as to what it says about his earlier work and the questions he was originally asked.

14. Mr Mookerjee's answer in paragraphs 2.59-2.60 is not relevant to this question. As stated, the denominators were vastly different between the RHC and the comparator hospitals. This question from the Inquiry specifically asks Mr Mookerjee for the entire analysis to be repeated with this second admission data set. He fails to do so.

15. In paragraph 2.62, Mr Mookerjee notes that, in comparing the first data set (overnight stays only) and the second data set (with day cases included), the admissions numbers are fairly consistent for wards 2A and 4B but have increased considerably for 2B and 6A. This is obviously to be expected given the nature of those wards. Wards 2A and 4B are predominantly inpatient wards where patients are admitted for overnight stays (often much longer). Similarly, the 6A data in the second data set is split into separate columns for 6A Inpatients and 6A Day Unit (Table 3 in the supplementary report), and the 6A Inpatients column is consistent with the first data set. Wards 2B and 6A Day Unit are predominantly the day case wards, so, when day case admission numbers are added, the numbers in the day case wards increase substantially. That Mr Mookerjee did not recognise this calls into question his understanding of the RHC/QEUH, the evidence before this Inquiry, and his entire analysis and conclusions.

16. In paragraph 2.63, Mr Mookerjee states that, rather than repeating his analysis with the full second admissions data set, he has restricted this updated analysis to 'wards 2A and subsequently post decant to 6A', though he neglects to specify that he is also excluding the column called '6A Day Unit', as is clear from his restricted data table on the next page (Table 4 following 2.65), which shows that only 6A Inpatient data has been used.

17. In short, having been informed that day cases were not included in the first data set, and having then been supplied with a data set that included day cases, Mr Mookerjee excluded the day cases from GGC's admissions data while keeping the day cases in the comparator hospitals. His only justification is that 2A and (inpatient) 6A are the only wards 'where we have utmost confidence that Schiehallion paediatric haemato-oncology patients resided'. This displays a fundamental misunderstanding of the RHC. It is not disputed that Ward 2B is the Schiehallion day case ward and that Ward 6A (Day Unit) was the Schiehallion day case ward post-decant. No witness suggests otherwise. These day admissions numbers do not include patients other than the paediatric haemato oncology cohort. There is no evidence to the effect that this is not correct. The admissions to the wards are clear and not disputed. Mr Mookerjee ignores this evidence and, instead, discounts data on an entirely flawed evidential basis.

11. Mr Mookerjee should carry out an analysis of the rates of gram-negative infections on a monthly basis from the opening of the new Schiehallion Unit in 2015 and 2022 first in Ward 2A only for those who stayed overnight there and then after the decant for Ward 6A both for those who stayed overnight only and those wo were admitted to Ward 6A irrespective of whether they stayed overnight. He should report the rates in each month, any trend in those rates, any observations he has about correlation, connection or association between those rates and both water testing results in those wards and interventions.

18. The Inquiry specifically asked for this analysis to be carried out on a monthly basis, and that Mr Mookerjee should report the rates in each month. Mr Mookerjee has not done so. The Inquiry also specifically asked that this analysis be restricted to gram-negative infections. Mr Mookerjee has not done so. The similarity of his second set of computed values to the first set makes it clear that fungal infections were still included. The Inquiry specifically asked that for 6A post decant, two analyses should be carried out (with rates reported per month, not per year): first, for those who stayed overnight only and secondly for all those admitted to 6A regardless of whether they stayed overnight. Mr Mookerjee has not done so.

19. The answer provided by Mr Mookerjee to the Inquiry's request is plainly flawed. He has explicitly excluded RHC day case admissions while keeping the day case admissions for the comparator hospitals. This is not valid. What he presents in paragraphs 2.65 to 2.67 does not address the Inquiry's request and is copied almost verbatim from his initial report. He claims to have recalculated the coefficient statistic but still fails to show any measure of confidence or hypothesis testing that would indicate whether this supposed correlation was likely to have arisen due to chance alone. These are fundamental and obvious failings which call into question Mr Mookerjee's conclusions and his reliability as a witness.

13. Given the changes in the data set Mr Mookerjee should provide his opinion and detailed explanations as what conclusions (if any) can be drawn from differences between the results of his earlier analysis and his later analysis…

20. Mr Mookerjee has not addressed this question directly. Instead, whilst it is not clear, he appears to have considered this along with Question 11. Given that Mr Mookerjee has not, in fact, conducted a new analysis that included RHC day case admissions data, and that he has copied over the work from his initial report, it is not surprising that he did not find any differences.

Comments regarding Mr Mookerjee's answers to the Direction 5 Questions

21. Questions from core participants were consolidated into fifteen questions outlined in the Direction 5 Questions for Independent Expert Witnesses sent to Mr Mookerjee. The questions are paraphrased below in bold along with NHSGGC's comments on Mr Mookerjee's response. These responses focus on the fundamental issues summarised above. However, these fundamental errors call into question the entirety of Mr Mookerjee's analysis.

Question 1 [12]. NHSGGC questioned why Mr Mookerjee included yeasts in his amalgamated blood culture infection rate, given that the majority of yeast infections in haemato-oncology patients arise from their commensal flora and not from the environment.

22. Mr Mookerjee's response at paragraph 1.1 does not address this question. He refers the Inquiry to paragraph 5.1.7 of his initial report, which does not exist, but if he meant to refer to 8.1.7, this paragraph mentions only bacteria, not fungi. It does not explain why yeasts were included. His answer to a question about the inclusion of yeasts, e.g. that 'comparing the Schiehallion to other large Trusts over a set period of eight years allows for a high level of confidence in the representativeness of the dataset and the outcome derived', does not make sense.

Question 1 [13]. GGC questioned Mr Mookerjee's correlation approach, given that the dominant bacterial and fungal species detected in water are rarely or never implicated in blood stream infections, and the dominant bacterial and yeast species implicated in blood stream infections are rarely or never found in water samples.

23. Mr Mookerjee's response at paragraph 1.2 does not address this question. He simply refers the Inquiry to paragraphs in his initial and supplementary reports that do not provide the required answer. He was not asked what he did, but rather why. Mr Mookerjee has not answered this or provided any further discussion of why it is appropriate to correlate agglomerated rates when there is minimal overlap in the bacterial and fungal species found in the two sets of data.

Question 1 [14]. GGC questioned why Mr Mookerjee included only Legionella spp., Pseudomonas spp., Cupriavidus spp., Serratia spp., Stenotrophomonas spp. and fungi in the water data. Several of these are extremely rare or absent from the blood stream infection data, Serratia is extremely rare in the water data, and several bacterial species that did occur in the blood stream infection data were present in the water data but inexplicably excluded.

24. Mr Mookerjee does not address this question. He refers the Inquiry to his answer to [13] at paragraph 1.2, which refers to various methodological paragraphs in his reports. As with [13], this does address the question that was posed to him. His selection of species from the large water data set that was provided to him is nonsensical.

Question 1 [15-16-17]. GGC questioned whether computing a single annual 'water positivity' rate and comparing its trend over time is appropriate given how water testing changed over the period under question, particularly with the addition of fungi- and Gram-negative bacteria-specific tests in 2018. Mr Mookerjee's approach assumes that the same 'thing' is being counted throughout this period and therefore that the values are comparable, which is not the case.

25. Mr Mookerjee's response does not address this question and is, in any event, entirely flawed. It is abundantly clear what he did to calculate water positivity. This was not NHSGGC's point. NHSGGC questioned whether this approach was suitable given the changes in water testing

that occurred over this period. It appears that Mr Mookerjee does not understand, at a fundamental level, how increasing the number of organisms that you test for might results in a wider range of organisms being detected. His approach to agglomerating the water testing data into a single annual 'water positivity rate' is not valid.

26. A simple analogy would be counting whether any vehicles were seen on a quiet road per day and calculating an annual rate (number of days with a vehicle divided by the total number of days), but in years 1 and 2 you count only white vans and only go out to count them one day per week, then in year 3 you count all cars and vans, and go out to count them every day, and in years 4 and 5 you also include motorcycles and pedal bikes. The number and types of 'things' that have been counted and grouped into a single 'positive' value changed so the 'rate of things' (e.g. percent of days with a vehicle) should not be compared over this time period. This clearly calls into question Mr Mookerjee's analysis and expertise.

Question 1 [18]. Given that Mr Mookerjee states that his computed water positivity rates for 2015 and 2016 are likely to be underestimates, GGC questioned whether it was appropriate to include these in any kind of trend analysis, and to conclude that there was a 'rising trend' of water positivity.

27. Mr Mookerjee refers to paragraphs 2.42-2.44 of his Supplementary Report, which he claims addressed Question 10i of the CTI Note. His answer does not adequately address that question. This is addressed above in respect of 10(i).

Question 2 [20]. GGC questioned whether correlation analysis was suitable, given that both agglomerated values were measured repeatedly over time and given the small number of data points involved.

28. In his response at paragraph 2.2, Mr Mookerjee appears to have misunderstood this question. The question was not about the 'rigour' of the underlying data or about changes in sampling methodology. It was an attempt to identify whether Mr Mookerjee was aware that correlation analysis is not appropriate for pairs of variables where each variable is time-constrained, i.e. measured repeatedly over time, and also that correlation analysis is of limited use with so few data points, as high correlation coefficients are likely to occur by chance alone (see discussion above regarding Mr Mookerjee's response to the question about small numbers of data points, 10e). These are not matters of opinion, nor are they advanced statistical concepts. These are fundamental features of correlation analysis that are covered in any introductory statistics textbooks, including the textbook Mr Mookerjee cites in the Initial Report. Mr Mookerjee's failure to answer these basic questions, and his insistence that a correlation analysis is valid here, calls into question his analysis and expertise.

Question 2 [21-22].

29. NHSGGC replicated the correlation analysis outlined in Mr Mookerjee's initial report, using his data table of agglomerated BSI and water positivity rates, and observed that a) plotting these values shows no clear linear relationship, b) the data points do not obviously cluster around a linear trend line, c) the correlation coefficient without the 2020 data is that obtained by Mr Mookerjee (though he rounds 0.66 up to 0.7 in his report), d) the correlation coefficient is lower when the 2020 data point is included (0.54), e) confidence interval calculation, which Mr Mookerjee omitted, show that the interval is wide and crosses zero, meaning that we cannot confidently state that any correlation exists, and f) hypothesis testing, which Mr Mookerjee omitted, shows that the p-value of the correlation that Mr Mookerjee carried out is 0.22 (or 0.26 if the 2020 data point is included). In short, Mr Mookerjee's own correlation analysis shows the opposite of what he has concluded: that there is no evidence of correlation between his computed BSI and water positivity rates.

30. Mr Mookerjee's response at paragraph 2.3 does not address any of the issues raised in question 21. He simply repeats his assertion that his reported correlation coefficient of 0.7 shows a 'moderate – very strong' correlation. He refers to paragraphs in the Supplementary Report that also do not address any of the points raised in this question. In particular, Mr Mookerjee does not address the issue of hypothesis testing and p-values, either in paragraph 2.3 or 2.4. He has shown no evidence that he understands correlation coefficients and why robust correlation analyses must always include an estimation of how likely it is that the reported correlation coefficient would be obtained through random chance alone (i.e. with uncorrelated data). This is what is expressed with a p-value – the likelihood (on a scale of 0 to 1) of obtaining a statistic at least as extreme as what was obtained (i.e. a correlation coefficient of 0.7, in Mr Mookerjee's report). A p-value of 0.22 means that with random, uncorrelated data, Mr Mookerjee's analysis would have generated a correlation coefficient of 0.7 or higher 22% of the time, through random chance alone, if carried out repeatedly. A generally accepted threshold for deeming a statistic 'significant' is 0.05, i.e. that there is only a 5% chance that a value that extreme would occur through chance alone.

31. His explanation at paragraph 2.4 about why he opted not to compute confidence intervals is incorrect and suggests a complete misunderstanding of the concept of samples and populations. He claims that 'since we are utilising real world data […] and not intending on using either statistic to infer something about a larger population, there is no need for confidence intervals around the statistics I have calculated'. This is plainly flawed. Mr Mookerjee does not have data for every possible water sample that could have been collected from the Schiehallion wards over this period (i.e. the population, in statistical terms). What he has is a sample. Had we gone back in time and collected different water samples on different

days, the positivity rate would be slightly different, due to random chance alone. His correlation analysis attempts to determine whether these two variables are associated, and a confidence interval around the correlation coefficient would give the range most likely to contain the 'true' value. He has given no indication that he understands this basic concept.

Question 3. The importance of significance testing was also highlighted by NHS NSS, who specifically asked 'Can Mr Mookerjee please explain what significance testing was undertaken?'.

32. Not only does Mr Mookerjee not answer this question at all (beyond referring to a mis-numbered paragraph), but he has shown no evidence that he understands what significance testing is and why it is the most important component of the type of analysis that he claims to have carried out.

Conclusion

33. For the foregoing reasons, NHSGGC submits that the Supplementary Report does not address the fundamental comments set out in NHSGGC's response to the Initial Report. On that basis, NHSGGC's comments and questions set out in its response dated 19 June 2024 still apply.

34. Mr Mookerjee's analyses and conclusions are fundamentally flawed which calls into question his credentials as an expert witness for the Inquiry. NHSGGC have provided the credentials of its experts who have carried out a review of Mr Mookerjee's analysis in Appendix 1. NHSGGC can provide further detailed analysis, should the Inquiry desire it to do so. However, if independent analysis is desired by the Inquiry, NHSGGC urge the Inquiry to instruct an independent review of Mr Mookerjee's report. NHSGGC consider that, given the fundamental issues in Mr Mookerjee's report, any recommendations made by the Inquiry will be unsafe.

Peter Gray KC
Emma Toner, Advocate
Andrew McWhirter, Advocate

Counsel for NHSGGC

29 August 2024

Annex 1

Dr Kate Levin

- Cstat (Chartered Statistician), PhD Health Sciences, MSc Social Statistics, BSc Mathematics and Statistics, Registered Chartered Statistician with the Royal Statistical Society

Dr Dominique Chaput

- BSc (Honours in Biochemistry, Minor in Mathematics), Mount Allison University, Canada
- MSc (Environmental Change), University of Oxford
- DPhil (Microbial Ecology), University of Oxford
- Ten years of postdoctoral research experience in environmental microbiology laboratories in the USA (Smithsonian Institution, Washington DC) and UK (University of Exeter), three years in a clinical microbiology laboratory (Scottish Microbiology Reference Labs)

Scottish Hospitals Inquiry

NHS National Services Scotland response to
Supplementary report by Mr Sid Mookerjee: Response to CP comments on his report 'Quantitative analysis undertaken to understand the association between the built environment and rates of gram-negative and fungal bloodstream infections at the Schiehallion unit between the years 2015 and 2022.'

1. In this response, NHS National Services Scotland (NSS) responds to the supplementary report by Mr Sid Mookerjee submitted by him on 7 August 2024. The comments below do not seek to raise new issues. They seek to clarify some aspects of Mr Mookerjee's methodology in respect of which NSS considers there still remain outstanding issues, in the hope that this may assist in properly understanding his conclusions. NSS considers that there remain significant questions over key aspects of both the principal and the supplementary reports. Due to the limited data available to Mr Mookerjee, which limitations he recognises, NSS does not consider that the strength of the conclusions which he reaches is justified.

2. At 2.2 to 2.4 Mr Mookerjee addresses the question posed at 2.1, to explain his decision to include certain infections and to exclude others in his Quantitative Analysis report. However, the episode case definition appears to be at species level. It is not explained why some key organisms do not feature in table 8.1.16 of the Quantitative Analysis report.

3. At 2.5 questions are posed about Mr Mookerjee's comparisons between the QEUH/RCH and comparator hospitals. Mr Mookerjee addresses these at 2.6 to 2.13.
    a) It remains unclear whether outpatients have been included in the denominator data. It is also unclear whether comparator organisations have included outpatient contacts and whether the denominators are comparable. This is not answered by what is said at 2.7.
    b) It remains unclear whether adult admissions have been excluded from the ward 6A and ward 4B denominator. This is not answered by what is said at 2.7 or 2.23.

4. At 2.14 Mr Mookerjee is asked to respond to NSS's response at para. 9 and Question B to the Quantitative Analysis report. He addresses para. 9 at 2.15 to 2.18, and Question B at 2.19 to 2.24. At 2.16 he continues to justify the use of the admissions denominator, as patients have contact during outpatient, inpatient and day admissions. It remains unclear whether outpatients have in fact been included in the denominator data, which may be contradicted by the definitions referred to in 2.12.

5. At 2.24 Mr Mookerjee refers to the caveats and nuances regarding the NHS GGC dataset raised in the NSS response, and states that they also apply to other large institutions in the UK. NSS accepts that this is likely. However, in epidemiological studies, differences across comparators and the effects of confounding are important to acknowledge. It cannot be assumed that the nuances in NHS GGC will be identical across the comparators, and this may impact on interpretation of the data.

6. At 2.41 Mr Mookerjee seeks to address NSS's point regarding the correlation analysis, however, his justification for including/excluding water sampling results remains questionable. Specifically, the inclusion of 2015 and 2016 results (of a small number) and the exclusion of 2020 results (of a high number) is not justified and undermines the conclusion of a strong association between water positivity and infection rates.

7. At 2.58 Mr Mookerjee is asked to use the Second Admission Data Set to repeat his analysis and draw conclusions. He addresses this at 2.59 to 2.63. NSS notes that this dataset includes the day case admissions and admissions for other patients who did not stay overnight. As a result, this dataset is better aligned with the comparator organisations' admission denominator. However, the analysis was not repeated using this entire new dataset, instead a subset of wards 2A and 6A was used (at 2.63). This subset also appears not to include day cases and appears to be virtually the same data used as in the principal report.

NHS National Services Scotland

2 September 2024

SCOTTISH HOSPITALS INQUIRY

Bundle of documents for Oral hearings commencing from 19 August 2024 in relation to the Queen Elizabeth University Hospital and the Royal Hospital for Children, Glasgow

Bundle 21 - Volume 7

Substantive Core Participant responses to Supplementary Expert Report of Sid Mookerjee