SCOTTISH HOSPITALS INQUIRY

Hearings Commencing
19 August 2025

Day 5
Tuesday, 26 August 2025
Mr Siddharth Mookerjee

CONTENTS

10:02

THE CHAIR: Good morning. Before inviting Mr Mackintosh to lead his witness for today, who will be Mr Mookerjee, can I just return to a topic I raised last week in relation to particularly the timetable of the submission of closing statements and the oral hearing we provided for next year?

As I indicated last week, having considered the impact of my proposal on a legal representative's professional commitments and entirely accepted that these should not be disturbed, the mechanism that I have adopted is to withdraw Direction 11. The document doing that will be designated Direction 12.

This should be issued this week, possibly even today, and effectively what it does is reinstate the previous timetable set out in Direction 10 and repeat the other provisions of Direction 11, so I trust this addresses the issue that arose. Now, Mr Mackintosh.

MR MACKINTOSH: Mr Mookerjee, please, my Lord.

THE CHAIR: Yes. Good morning, Mr Mookerjee. You have, of course, previously given evidence to the Inquiry and will be broadly familiar with our procedure, which has not changed since you were last here. But, as a preliminary, I understand you're prepared to take the oath.

THE WITNESS: Yes, I am.

Mr SIDDHARTH MOOKERJEE

Sworn

THE CHAIR: Right, thank you, Mr Mookerjee.

THE WITNESS: Thank you.

THE CHAIR: Now, as I'm sure I said to you on the last occasion, it's important that we hear what you have to say, so maybe speak a little----

THE WITNESS: Yes.

THE CHAIR: -- more loudly than you normally would. We anticipate that your evidence certainly will go into the afternoon, but we will take a coffee break probably at about half past eleven. However, should you wish to take a break at any stage, please just give me an indication and we can do that.

THE WITNESS: Sure.

THE CHAIR: Now, Mr Mackintosh.

MR MACKINTOSH: Thank you, my Lord.

Questioned by Mr MACKINTOSH

Q Mr Mookerjee, I wonder if you can confirm your full name.

A Yes, so, morning, I'm Sid Mookerjee.

Q What's your current employment?

A So I work as a hospital epidemiologist within the Infection Control department at University Hospitals Sussex, and I am assistant professor in global health and infection epidemiology at the Brighton and Sussex School of Medicine.

Q Thank you. Now, you gave evidence last year on 5 November, and I'm proposing to ask you questions about epidemiological matters that have arisen since then, and they fall into two groups.

Now, the first, which we'll deal with first, is a critique of your comparison exercise contained in your reports in bundle 21, volume 1, and secondly the work you did for the Inquiry in responding and reviewing the HAD report, which we can find in bundle 44 and its various volumes. Now, I'm proposing to do that after the first one.

What I wanted to do first is to take you to your original report, so that's bundle 21, volume 1, document 1, page 19. So we land in the early stages of this report where your discussion is-- what you aimed to in in the original report is at 6.2.

A Yes.

Q The second bullet point, what was that involving you planning to do in your original exercise?

A So with regards to .2 and 6.2, the aim was to look at the infection episodes as a link to environmental bacteria and fungi in comparator institutions which were large enough and comparable enough to the RHC for the period from '15 to '22.

Q Now, I think you're aware of criticisms that we went through when you last gave evidence of your comparator exercise, largely raised by NSS ARHAI, that you had failed to take account of confounding in the design of this study. Just to put this in context, can you set out what you understand the criticisms were and how you would respond to them, just in summary terms?

A So I guess, in summary terms, the criticisms were-- well, the first one, as I understand it, was that were these comparator institutions ones that you can-- you could compare the RHC to? So, from that, I got that the criticism was that-- was the paediatric haematology unit and the patients in between the RHC and its comparative units in a state that they were comparable?

And the second was about, are we comparing like to like in terms of the methodology of-- in terms of how the infection rate has been calculated?

Q Was there a particular aspect of that second criticism which was related to the method of deduplication?

A Yes, so to expand on the second one, the point was raised, or the question was asked, whether the comparator institutions abided by what was asked of them, which was that we need deduplicated infection episodes.

Q When you communicated with the comparator institutions in a Freedom of Information request, did you tell them what deduplication standard to apply?

A So the Freedom of Information stated explicitly that we wanted the infection episodes to be deduplicated.

Q But did it say how to do that, the methodological----

A It did not, from what I recall, explicitly state the methodology that they should adopt, but I think the expectation was that, since large acute trusts, as part of their routine and monthly reporting of infections to-- for example, in England to UKHSA and in Scotland to the corresponding entity, submit deduplicated infections on a monthly basis, as is mandatory to these institutions, that that method, which is recognised as the only-- well, you know, or is recognised and is consistently applied in terms of the 14-day rule to bacteraemias, would be applied to the infections that they sent back.

Q So just before we go into that detail and come back to your response to the criticisms, you mentioned how hospitals in England might behave in respect to monthly reporting data. You also mentioned Scotland. Do you have any knowledge about whether Scottish units send monthly bloodstream infection results to deduplicated to a national agency? Do you know whether that happens in Scotland?

A Yes.

Q Right, but if we focus on the English units, effectively you didn't tell them how to deduplicate because you assumed they were going to do it the way everyone else does it in England.

A Absolutely.

Q Right. Let's go back to the first criticism, the one about the comparability. How did you address the criticisms from, amongst others, NSS that you weren't taking account of confounders in these differences between what turned out to be two small units, Cardiff and the Vale and Oxford, a larger one in Leeds and a much larger one, albeit not the same age range, in Great Ormond Street? How did you take account of that in your methodology?

A So in epidemiology, it is recognised that one crucial way in which you can adjust for confounders, which you accept are at play in any large NHS institution, is to make sure that the data that you are using to say something about them, so in this case infection

rates, is for a period that is large.

And by that I mean that the range, in this case years, is a long one. Secondly, that when you are using the data whereby you're comparing, in this case, as we did, the RHC to the comparator units, that you make sure that your assumptions are based on a lot of data, so a considerable number of infections and a considerable number of activity data.

Q So if we come back to the number of infections----

A Sure.

Q -- what's the size comparison between the data set for Glasgow you were looking at in terms of the number of infections in broad terms you found, which seem to be in the high hundred-and-somethings, and the number of infections you found across all the comparator units? I mean, how do the two numbers compare?

A So in terms of approximate comparisons, yes, so we were comparing 187 environmental organisms at the RHC over 2015 to 2022, as compared to approximately 3,000 or so environmental organisms for that period from the comparator units.

Q For the four comparator units?

A For the four comparator units.

Q When you look at the activity there-- I don't want to reopen the discussion about, is admissions better than bed days. I think we've done that, and we move on, but just in terms of-- I think you eventually used admissions, what's the differential between the number of admissions effectively in the Glasgow Schiehallion activity measure and the scale of the admissions across the comparator units?

A So the comparator units approximately had 150,000 admissions for the period 2015 to 2022, and the RHC would probably be around 10,000 to 12,000 or maybe in the ballpark of that.

Q So putting aside the deduplication criticism for a moment, to what extent are you effectively taking the position that the scale solution addresses all the confounding problems? Is that effectively your response?

A Well, absolutely, because in epidemiology what you want is-- or what you're saying is, if the scale of the data is large enough, that the precision of the estimate is more precise, so the larger the sample data, the more precise the estimate.

Q Okay. If you go back to bundle 21, volume 1, page 19, in section 7, you have a table which is split over a page, 7.1. It makes it slightly harder to read, but if you look at the first box, and this is not the whole of the first box, the data is "Patient infection episodes". The data

specific then says “QEUH including RHC”. if we go over the page and we go back again, the case definition is, “Gram negative bacteria – environmental [over the page] and enteric group and fungi”.

Now, if we then go back over the page to the first page, to 19, and then you have the details of the analysis,

> “Gram-negative and fungal bloodstream infection positives for the [over the page] ‘Schiehallion patient cohort’, i.e. samples taken in wards 2A, 2B, 6A, 4B for the period 2015-2022.”

Do you effectively list the infections that fall within that case definition elsewhere in the report?

A So the question was, do I list the organisms?

Q The organisms and the numbers elsewhere in this report----

A What I----

Q -- that fall within that case definition.

A So, the answer to that is no, not explicitly. What I do make note of----

THE CHAIR: Sorry, “No, not exclusively”? Sorry.

A No, not explicitly.

THE CHAIR: Explicitly?

A Yeah, explicitly.

THE CHAIR: Thank you.

A Yeah, sorry, my Lord.

MR MACKINTOSH: So because if we go to page 25, what is this showing?

A So, page 25?

Q Yes.

A Yeah, 8.1.16 is the summary table of the environmental bacteria and fungi which were attributable as per the spreadsheet that was provided by NHSGGC to the physical spaces that were 2A, 2B, 4B, 6A, so it is essentially the final list of organisms that I arrived at after my analysis.

Q Is it effectively, if we go back to page 19, the product of the process that starts with that case definition?

A Yes, absolutely, and I recognise that, you know, in hindsight that should’ve been made very clear, but yes, the start of the process was the case definitions which are here.

And we then moved through the iterative process of ensuring that we take into account all the environmental infections leading on from both bacteria and fungi for the RHC there where we were focusing on 2A, 2B, 4B, and 6A, and then-- so doing that for the comparator institutions.

Q So if you take the case definition of the first row of Table 7.1 and you put it through your process that you did last year, your position is you get to the table at 8.1.16, page 25?

A Yes, yes, yeah.

Q Right. If we go back to page

20 and we look at the comparator institutions row, you have the same words in the third column, “Gram Negative bacteria and enteric group and Fungi.

A Yes.

Q The source is FOI, and then you have-- If you jump back to the previous page, is the difference of the definition limited to just geography, back to page 20, or is there any other difference in the definition?

A No, the difference of the definition in terms of the details of the analysis is only-- the difference is only that, for the Schiehallion patient cohort, we took the physical spaces of 2B, 6A and 4B, and we took everything that was environmental bacteria or fungi from the other units.

Q But in terms of geography, you gave them the definition of the paediatric haemato-oncology unit and allowed them to work out what was included?

A Absolutely, yeah, that’s what we did.

Q Now, is there in your original report and any of your reports to this date a list of the species and genesis of gram-negative environmental and enteric group and fungi that appear in that third column? Have you got a master list in your reports so you can sort of check off against the master list?

A No, I don’t think I included a master list because I-- Saying that, I think there is a broad list that I include in one of my 2025 reports, where I compare the organisms of-- that are included within the CNR and compare it to the HPS and compare it to my, well, quantitative report and the HAD report. But in answer to your question, well, prior to that, there is no master list.

Q Now, what I want to just check is two things. They’re quite long questions, and so I’m going to give them at a slow speed. If there was a species of bacteria that caused a bloodstream infection in the Schiehallion unit that met that case definition in the third column on page 20, would a bloodstream infection of the same species in any of the comparator units be counted for the comparator units?

A Absolutely.

Q Yes. If there was a species of bacteria that met the case definition that didn’t cause an infection in the Schiehallion unit but did cause an infection in one of the comparator units, would that be counted in their rates?

A Absolutely.

Q Right. Would you accept the criticism that this is not clear from your original reports?

A Yes, I do accept that.

Q Now, you produced your

report. You went through a series of iterations, one of which involved a replacement of your admissions data set by a new admissions data set.

A Yes.

Q Can you explain, from your perspective, how that came to be necessary, the replacement of a first admissions data set for Schiehallion with a second one?

A So, as I recall, once a near draft of the-- or it might have been the finalised-- the quantitative infection link report written by me was circulated, that a-- that a response was received from NHSGGC through the-- through the SHI, acknowledging that the original admissions data sheet excluded patients with a length of stay less than one, which essentially meant that they had excluded from the admissions numbers anyone with a day stay.

So they would have come in-- which is very well often in the case for paediatric for haematology. The patients, they had come in to the hospital but they hadn't stayed overnight. That-- they had understood that that had needed to be corrected and they responded by issuing a amended admissions sheet, this time giving the number of admissions but including the admissions of patients where the length of stay was less than one.

Q Would you accept that they did actually say that on their document, you just hadn't noticed?

A Yes, absolutely. So, it wasn't something that I was aware of, and I was only well made aware of this once they had seen the analysis that I had issued through the quantitative report.

Q Then it became necessary to do a secondary calculation, didn't it, just----

A Yes.

Q -- before the hearing? What caused that?

A If I remember correctly, the second recalculation right before the last time I gave evidence was in light of the point raised that my rate of infections had, so until that point, been based on admissions----

Q Yes.

A -- and a rate per thousand admissions, and that bed days is an equally important activity measure, and so, if I am-- if I remember correctly and do remind me, that I then calculated a rate of infection. Now, it may have just been for 2A.

Q Could it be that you had some concern about mapping patients to wards at this point?

A Yes, so that has always-- or that I should-- yeah, I should note has been a concern from the beginning,

because the way in which we understood the-- or the way in which we went about----

Q So, in this case, it's you and Dr Mumford and Ms Dempster?

A Yes, absolutely. Went about understanding what the rate of infection within the Schiehallion cohort to be is that we-- we looked at the physical spaces of 2A, 2B, 4B, 6A to be the physical spaces that encapsulate the Schiehallion patients.

It was-- so, and therefore, in terms of the infection data for completeness was based on samples that were linked to the physical spaces of these four wards.

Q Does that mean that on the sample record it said 2A?

A Yeah, absolutely. That----

Q If it said another ward, you wouldn't count it?

A Yes, would not be counted. So it had to-- you know, it had to explicitly say that this sample was taken on 2A, 2B, 4B, or 6A. Now, the difficulty was, of course, to emulate that for the admission data, and that took a lot of work, a lot of work between, so, myself with Dr Mumford and with Ms Dempster to map as accurately as we could----

Q Yes.

A -- the admissions that were peculiar to 2A, 2B, 4B and 6A. So when the second admissions sheet was issued, that whole process had to be redone. So that we could-- we, again, had to calculate from scratch what the admissions were peculiar to 2A, 2B, 4B, 6A would be.

Q Right, what I want to do before I turn to the criticisms made of you is to sort of get you to the point you got to when you gave evidence. So, we can look at bundle 27, volume 17. I'll just try and remember the page number. No, volume 18, sorry. It's page 3. Yes. So you talked of this chart when you last gave evidence?

A Yes.

Q What I want to do is just to get you to re-explain what all the elements relate to. I'm not going to ask you to reinterpret it because you've done that already.

A Sure.

Q Just to provide context of where we're going, I want to walk through each bit of colour on here and work out what they are. So, the blue vertical bars, what do they record?

A So, the blue vertical bars is a percentage figure for water samples which were positive in that year.

Q Now, did you eventually do a comparison between the water positivity rate and one of the other lines on this chart?

A Yes.

Q Which was the chart you compared the water positivity rate with-- the line you compared with?

A So, I compared the water positivity rate to the purple line, which is----

Q The bottom purple line?

A -- the overall Schiehallion rate per thousand admissions.

Q That's the one that has a peak at 25.7 in 2017?

A Yes.

Q Right. Now, this is probably a good moment to put to you a criticism made of that statistical comparison by Ms Cairns of NSS. She, I think, acknowledged that there were fundamental issues with the availability of water testing data, particularly in ’15, ’16, and ’17----

A Yes.

Q -- but explained that in her view, doing a statistical comparison between annual water positivity results and any form of infection rate won't work because there are simply too few data points here, partly because there's too few test results, but partly because you're only doing it annually. How would you respond to that criticism?

A I think I would accept that criticism. It was a facet of reality that the number of water samples that were taken, so initially in 15-16, were a lot fewer than the ones that then were taken, for example, comparing it to ’17, ’18 and ’19. Therefore, the water positivity rate is based on fewer samples which were taken and fewer samples which were positive of those samples.

So I take that criticism on board, but I just-- I would like to caveat that by saying that that was all the data that was available. So, it was a facet of-- yeah, so what had happened. The second thing I would say is that what we eventually did, and you can see from this graph here, is that we took the water-- we took the rate of water positivity in these blue bars over, so, 2015 to 2020----

Q Yes.

A -- and when we say we compared the water positivity rate, what we compared really in epidemiological terms was the trend in water positivity over the time period, so 2015 to 2020, to the trend of the overall-- to the overall Schiehallion rate per thousand admissions.

The trend was analysed using a tool in-- tool in epidemiology called a correlation coefficient which essentially, so, tells you whether, if you lay the two trends side by side, how well do the dips and the peaks and when these peaks happen and when the dips happen, how well do they correlate with each

other.

So the correlation coefficient tool, well, gives you a value and the value is a measure of concordance and that value that I calculated would have been a-- So value that is a measure of how the overall trend of infections aligns itself to the overall trend of water positivity.

I got a high-ish value of, I think, and I can be reminded of it-- well, maybe if we go back to the quantitative report, but I think it was around 0.7, if I'm not wrong, or thereabouts, which, from the literature, what you can take from that value for the correlation coefficient is that there is a strong to very strong association or concordance in the trend of the Schiehallion rate per thousand admissions with the overall trend of positivity from the water samples.

Q Could that concordance be driven actually simply by the quantity of testing, particularly in 2018? So 2018 is simply distorting the whole exercise by being this massive peak of testing in, I think, February, March, which we see in Dr Chaput's reports?

Could that just sort of effectively skewer the whole thing in one direction and, therefore, Ms Cain's criticism has some value because whilst you've used what you can use, what you have available to include is this massive peak of testing?

A Yes, I think I would accept that. I don't think we can-- I think we have to accept the argument that water samples were taken as a reactive measure and therefore more water samples were taken for a period in reaction to what was being seen as rates of infection.

Q Now----

THE CHAIR: The perhaps simple point, and I apologise for making such a simple point, the more water tests you carry out, the more samples you take, the likelihood is you're going to find more positives in raw numbers?

A In raw numbers, yes, on the proviso that the rate of contamination of the system remains the same. Because, essentially-- so, if everything else remains the same, the more water samples you take, you would expect a higher proportion-- well, not a higher proportion but just more in terms of the numbers of water sample positives. What I would like to say here is that that per cent or per hundred is what I have here.

THE CHAIR: Okay.

A So what the blue bars here are or what we can take from the blue bars here is per cent or per hundred, what is the rate of positivity. Now, with all of those caveats in terms of reactive water testing, what it does-- will tell you, this

graph, is that 7.6 per cent of samples were positive in 2017, 17.36 per cent of all water samples were positive in 2018, and 11 per cent in 2019, 6 per cent in 2020. So what it says is that a higher proportion of the water samples taken in 2018-19 were positive as compared to 2017. I won't compare it to 2015 and to 2016 because we recognise that the context was not the same then.

MR MACKINTOSH: If we go back to this chart on page 3, I think the question about the green line, the yellow line, and the dotted purple line, do these actually come from previous calculations before you adjusted your admission day data in an earlier report? This is about a chart taken from an earlier report, and so these three lines, the green, the yellow and the spotted purple, are actually the chart, the lines out of your supplementary report?

A Yes, these lines are from the supplementary report.

Q Therefore, they predate one of the steps of realisation about the nature of the admission data?

A I don't think that's the case.

Q You think they postdate the realisation that----

A They postdate the realisation.

Q That the admission data was overnight only?

A That the-- yeah, that the admission data was overnight only and therefore they are based on the second admission sheet, which takes into account the overnight stays.

Q But they might precede this anxiety that patients aren't necessarily where you expect them to be?

A Well, I think that that anxiety has been around for the entire time----

Q Right.

A -- and we took that anxiety into account when we specified, as you can see for this analysis, and went with the methodology that let us look at the ward areas where we are most certain that Schiehallion patients resided, which was 2A and thereafter from 2019 in 6A.

Q So that's where the dotted purple line comes, and that's the combination of the two?

A Yes, absolutely. So the green line is the standalone 2A infection rate line, the yellow is the standalone 6A infection rate line, and the dotted-- well, purple is-- well, one follows the other so that you have one line that you can go on.

Q Just to be clear, these are on admissions, not on bed day?

A Yes, these are on admissions and not on bed day.

Q So having just done some context, let's talk about the purple line and the dotted red line. So you've

described the purple line as the overall Schiehallion rate per a thousand admissions.

A Yes.

Q We see that calculated in the table below in the fourth column that starts 5.37.

A Yes.

Q Then the number of infections is in this third column starting with a seven, and that's what adds up to 187. Or does it not quite add up to that?

A Yes, so the column that reads "Cumulative infections" from Table 8.1.15 should add up to 187.

Q Okay.

A The overall-- I can be corrected on that. I mean, it's been a while since I wrote this report, but I think that----

Q But it is, in fact, the numbers that come out of bundle 21, volume 1, page 24, and that over the page?

A Yes.

Q Right.

A Yes, so these are----

Q If we go back-- Sorry, carry on.

A Yes, so yes, these are-- this table and the cumulative-- the column from this table, which carries on to the table that we were just at, should equal 187 and is the overall incidence of environmental infections at the Schiehallion 2A, 2B, 4B, 6A.

Q Okay. If we go back to volume 27, volume 18. That one, thank you. If we look at the overall comparator rate, so this is you calculating from the number of infections supplied by the four comparators through a process which we can describe and deal with in a moment, divided by the admissions provided by them.

A Yes.

Q Right. Now, if we take that off the screen, what I'm proposing to do is take you to a report-- well, a document produced, I think, last week, possibly a little bit before then, by Dr Chaput who gave evidence on Tuesday. So this is bundle 44. Well, sorry, I'll rephrase that. Sorry, that's not quite right.

I want to go to our original December 2024 report, so that's bundle 44, volume 4, page 3. Now, I just want to be clear. I think you've actually answered this question, but I feel it's important to be clear.

If we turn onto page 5, Dr Chaput described in her text that Table 2 contains a list of organisms contained in multiple sites, and the first column is organisms contained both at Schiehallion and elsewhere and therefore I'm not going to ask you about that because I think you've already been clear, and she's clear. Those are in your data set,

in your table in bundle 21, volume 1, page 25, but the right-hand column-- Were these species counted in the analysis of the comparator datasets to calculate infection rates for the comparator units in which they appeared?

A Yes.

Q If we go on to the next page, she's then analysed for five different sites, albeit not referred to them by their names, but for all the infections listed there for Site A, Site B, Site C and Site D that didn't occur at Schiehallion, were they counted within their appropriate infection rates in your original analysis?

A Yes.

Q Yes, okay. Would you accept that wasn't perhaps as clear as it could have been in the original report?

A Yes, I accept.

Q Okay. Now, what I want to do now is move to her latest document, which actually you've attached as an addendum to your report that we asked you to do, and so what I'm proposing to do is to get a series of documents on the screen and then work through them methodically.

A Sure.

Q So I wonder if you can go to bundle 44, volume 9, and we have a report from you, and we'll go to-- sorry, go back to the index page, and what we'll do is we'll ignore your report for a moment, but look at Dr Chaput's notes, which is on page 17. So if we could zoom out so we can see the whole page on the screen. Have you read this?

A Yes, I have.

Q Right. Go over to the next page. The way that I read Dr Chaput as structuring this, and I think it's clear from her evidence, that she's looked at each comparator unit and raised a concern about deduplication, but for two of them she's raised other concerns.

A Yes.

Q And then she's raised concerns about deduplication for the Glasgow dataset. What I'm proposing to do is to work through these slowly. Now, I also have other documents I need to refer to. I mean, take that off the screen for the moment. I have your response report.

Now, my Lord, we created a bundle of the Freedom of Information Act requests that were received from all the four comparative units, and this is bundle 44, volume 10. I'm not proposing to put this bundle on the website, my Lord, and my reason is this: it contains the infection rates for all the other hospitals. It's not of any interest to the Inquiry what the particular rate of infection was for a particular organism in a particular hospital, a particular year.

We're simply looking for the aggregate and so I don't intend to take Mr Mookerjee through the individual hospital results, but I'm also concerned that it's not really within our remit to be advertising infection rates for other hospitals historically when they provided us in order to construct this collective sector rate.

All the core participants have the bundle under Restriction Order 1, but it won't go on the website. I will have a look at a couple of the pages in it which don't have lists of infections – they do have admission numbers and things – just to deal with Dr Chaput's concerns.

So if we can have to hand 4410, but don't quite go to it yet. Now, what I want to do is go back to the page in Dr Chaput's document we just had on the screen, so that's 44. Thank you very much. So Cardiff and the Vale. Now, what do you understand the criticism being made of you by Dr Chaput to be in the section here on page 18?

A So, as I understand it, the anxiety by way of the opinion here is that the comparators of Cardiff and the Vale and Leeds did not deduplicate their infection samples when they sent it to us.

Q What do they say? Well, let's actually look at the response itself. So if we go to bundle 44, volume 10, at page 46.

So this is a two-page cover letter that appeared, which was sent to the Inquiry in response to our Freedom of Information request, from Cardiff and Vale University Hospital Board, and if we just step onto the second page, page 47, go back to page 46, we can see the questions that we asked and the answers that we got. Now, do you see the section that Dr Chaput has quoted at item 4, Mr Mookerjee?

A I do.

Q And they've said:

"We have attempted to deduplicate these samples, but we're unable to guarantee it's 100% accurate as patients can send multiple blood culture samples and we have multiple organisms from blood culture bottles. Therefore, the higher number of isolates compared with the number of positive bottles is likely where both bottles may have grown a organism. It is also possible that there may have been more than one organism from a positive bottle."

And then over the page, we've asked the main question:

"A list of the numbers of organisms by species isolated from blood cultures from patients on paediatric haematology-oncology unit, whether deemed significant or not, by site, peripheral venipuncture, peripheral line or central line, by year from 2015-2022, total and

deduplicating numbers for the same infection episode. Please see attached."

Now, what was your approach to the information that was attached to this in terms of deduplication?

A So, if you go back to page 46, with reference to the line firstly, so:

"Therefore, the higher number of isolates compared with the number of positive bottles is likely where both bottles may have grown an organism. It is also possible that there may have been more than one organism from a positive bottle."

This is not a unique statement in that in every lab in the world that is assigned to a hospital, you will have to deal with the fact that from the same patient you might have a lot of blood cultures, and that some of the blood cultures might have more than one organism and that you will have some blood cultures where the same organism is put twice. So I didn't see that as unique or especially, well, exciting. I then went-- We then go up to the paragraph right before:

"We have attempted to deduplicate these samples, but we're unable to guarantee it's 100% accurate…"

So I took that at face value that Cardiff and Vale have tried their level best to abide by the question, which is clearly-- which is clearly put, and they have to the best of their abilities deduplicated the samples, and so I took it on face value.

Q If they hadn't fully deduplicated the samples, which is to be fair what they almost say there, what would be the effect on the calculated incidence rate for Cardiff and the Vale? What would it do to it?

A It would have inflated it.

Q In the context of a comparison with a rate in the Schiehallion unit, would that increase the chance of there being a difference or reduce the chance of there being a difference?

A It would reduce the chance.

Q Why?

A It would reduce the chance because you would be comparing deduplicated episodes of infection, so aggregated into a rate for the Schiehallion for each of the years, to a un-deduplicated-- so essentially all environmental and fungal infections from Cardiff and the Vale. So the difference----

Q There would be more infections in Cardiff and the Vale recorded than there actually were?

A Yeah, exactly.

Q Right. If we go back to bundle 44 volume 9 and the page-- Yes, we look at the lead section below. So Dr Chaput's note says:

"The FOI return from Leeds clearly states the data have not been deduplicated at all contrary to what you asserted in your reports and oral evidence."

She is right. You did assert that it was deduplicated in your evidence.

A Yes.

Q Yes. If we go and look at the actual response, which is volume 10 page 52, jumping rather than going through the pages. So this is a letter to us and we look at the response. I think the detail will come on the next page. Now, question 6 I read out before, but what's the answer to question 6, Mr Mookerjee?

A So the answer to question 6 is – and I read here – what they have offered in that summary, in responding to question 6, is they've noted the way you request total and deduplicated organisms by episode, the Telepath, which is a system that is widely used by the NHS, does not carry data on what constitutes an episode. So we've been unable to provide that part.

From that-- So if you may, what I understand is that they were able to respond to question 6 by giving us the total and the deduplicated organisms, but they weren't able to deduplicate it by episode.

Q What experience do you have of using Telepath?

A It was the system in place when I joined the NHS Trust which was Imperial Trust in 2011.

Q So how many years did you use Telepath before it stopped being used or you left?

A I would have used it for two and a half years.

Q Now, if we think about this criticism, how do you respond to Dr Chaput's comments? If we go back to volume 9 on that page, 18. Do you accept what you're saying or do you do not accept it?

A So, I do not because what is being implied here is that you need data on the episode of admission for a patient to be able to deduplicate the blood culture list, wherein my experience – which now exceeds a total-- so of about 18 years – is that all you need is the collection date of the blood culture and a unique-- and a reference which is unique to a patient.

So, for example, in the spreadsheet that is the infection episode from GGC, each unique patient has a unique CHI number and it is with those two sets of data that essentially, in very simple terms, you can arrange the blood cultures from oldest to newest and apply the deduplication to the list.

Q So are you effectively saying

that just because Leeds doesn't retain episode data, it doesn't mean it couldn't deduplicate?

A Absolutely. I have never found to be in need of the episode-- so-- in order to deduplicate data, if I have the collection date of the sample and the hospital number or a-- you know, yeah.

Q If we can go to page 19. Now, if I understand, Dr Chaput's criticism here is that you've only had part of a year's data for 2016. You basically had October, November, December infection data but you've divided that by a whole year's admission data. Is that correct?

A So I think I-- So the look on my face is me trying to remember.

Q Would it help if we looked at the chart we reproduced?

A Yes, yes it would help.

Q So this is bundle 21, volume 1, document 3, page 86. This is Figure 2. Ignore the purple line at the top and we look at the lines at the bottom. What were you plotting on this chart?

A So this chart gives you the individual infection rate per 1,000 admissions by year for each of the four comparators. So you have the dark green line for Great Ormond Street, you have the purple line for Leeds, you have the sky blue line for Cardiff and the Vale, and then you have a lighter green for Oxford and then a mean or an overall rate by year, so taking into account all of those four comparators year on year.

Q I mean, what was the point of plotting this chart when you did it?

A The point of plotting this chart was to allow us to have a visual on the extent to which, if there is-- what's the word, if there is something to be said about the rates of infection for-- or the rates of infection between these four units to be different for any one year. So essentially, what I wanted to sort of say was to look at the variability or the spread in infection rates from these four units----

Q Are you effectively looking to see if one of these units is doing something unusual compared to its colleagues, as it were?

A Absolutely, so I wanted to visually be able to represent whether, yes, one of these units is an outlier or has a rate that is drastically not like the others, and what I took from this chart was that these rates look fairly similar to each other.

So if you look at the data points, for example, there is a equal-- there is a good spread of the data points above the line as they are-- so under the overall line, which is the one in dotted red, showing that it takes into account the fluctuations that you would see year on year from four units but that nothing stands out as being unique to any----

Q I suppose two points. One is the purple line has a very low point in 2016. Is this period which Dr Chaput has identified when you've got a third of a year's worth of admissions, so might it be that you divided in that data point a third of a year's infections divided by a year's admissions?

A So I think I accept that for one of these graphs that I may have taken into account the admissions and that may have unjustly deflated that 2016 rate for Leeds, but that would've been contained to just one rate----

Q In one year?

A In one year for one hospital.

Q Now, I want to just understand mathematically how you calculate the overall comparator rate by reference to this chart because I think there's a point of understanding which I want to get clear, certainly my mind. When you calculate an instance rate, we've learned a lot about numerators and denominators.

A Yes.

Q So to take, for example, the first data point on the Cardiff and the Vale line, you would've taken the number of infections you found that met the case definition in Cardiff and the Vale in 2015, divided by the number of admissions they gave in Cardiff and the Vale, and that gives you a data point on that chart?

A Yes.

Q Right, and that applies to all the data points on that chart?

A Yes.

Q How do you work out the data points for the overall rate? What's the mathematical process for that?

A So the mathematical process for that is that you would take into account all the infections in the numerator from all the four hospitals.

Q So you effectively add them together?

A Yeah, you add them together, yeah, exactly. And then you would divide that by the sum of all the activity----

Q All the admissions?

A All the admissions for those four hospitals, and then to get a per thousand you would then multiply it by a thousand so that you can get a per thousand rate.

Q So although we might think-- I mean, the two things possibly might be mathematically the same, but we might think you're averaging the four rates in a particular year. You're not, actually. You're just creating an aggregate rate.

A Yeah, absolutely. It's not an average, it's an aggregate.

Q As if there's one big hospital.

A Exactly, because that in epidemiological terms is what you're

trying to do, is you're trying to, for every-- If the aim is that we want to compare a rate that we are interested in and we want to compare it to something else, that something else for each of the points in time needs to be based on as big a numerator and as big as a activity as possible. So we take the aggregate because it gives us the maximum certainty that that estimate is a good one to compare to.

Q Okay, so we think about the criticisms that Dr Chaput has raised in respect of Leeds. If Leeds had not deduplicated, what effect would that have on the overall comparator rate?

A Well, first of all, if I may, if Leeds had not deduplicated, you would see that in this because Leeds would stand out as an outlier. So essentially, that purple line would be far above the trend lines for the other comparators.

Q Are you effectively taking some reassurance from the fact they're all in the same place?

A Absolutely.

Q Right, and I think we discussed the impact of the 2016 number, so I'll move on. Let's turn to Oxford, so back bundle 44, volume 9, page 20. Now, Dr Chaput has said, "No details were provided in the FOI return about how data were deduplicated." She suggests that the layout suggests that that might render it difficult.

Now, I don't want you to comment on the clip she's put in Figure 5. I'd prefer to look at the whole document, and so what I'm proposing to do is take you to bundle 44, volume 10, and we want to go to page 59. I wonder if we can just-- Yes.

Now, passing over the charming address, Oxford's Freedom of Information has referred to us as "Dear No Name Provided". I'm grateful for their clarity. What Oxford seems to have done is they've inserted some of the answers into the text, but these ones don't contain infection rates, so we'll look at this happily together. If we can go onto the next page, we get a total number of blood cultures on Question 4, and then we have Question 6, so the question we asked is:

"A list of the numbers of all organisms, by species, isolated from blood cultures and patients on the paediatric haemato-oncology unit (whether deemed significant or not), by site (peripheral venipuncture, peripheral line or central line), by year for 2015-2022, total and deduplicated numbers for the same infection episode."

They then say, "Answered in attachment." What was your conclusion about whether Oxford had deduplicated?

A What I took from this was that Oxford answered the question and that

they had provided us with the numbers that-- as we had requested. There was no reason for me to think anything else.

Q What about Dr Chaput's reason, which is on page 20 of bundle 44, volume 9, where she explains that the way they've reported the results but divided in some cases between red ports and white ports would indicate incomplete deduplication? Why would you not think that was a reason to be suspicious?

A Because I don't concur with her reasoning in the way that she is reading what was sent back. I see this as a list, well, provided by Oxford. And so if you note that FOI, when you send it in, first of all, you might get replies back, like, for example, from Leeds and from Cardiff and Vale, where they will, and they should and they are required to, enter or to summarise the caveats inherent within their answer. So Oxford has not done so.

I am sure that this isn't the first FOI that Oxford has had to answer. They've been around for a very long time. Most NHS trusts are dealing with at least anywhere between-- I mean, specific to infection episodes, in my time at NHS trusts, which now extends to 16 years, there's a minimum of three a month, so this wouldn't have been new and they had ample opportunity to ask SHI for clarification.

Q But they didn't provide any caveats, and that's good enough for you?

A Yes, they did not and that is based on the proviso that this is an open conversation being had through the FOI.

Q If they hadn't deduplicated, I mean, would the answer be the same as the Leeds one, that it would stand out in the data, in the chart we just looked at?

A Absolutely. It would stand out, and that was the reason, or that is the reason, why that chart is so useful, because it provides you with what is essential in, you know, the first few steps of investigating from an epidemiological perspective anything, is that you lay out the data and that you visualise it. And there's a lot to be said to be able to do that and to infer things from that visualisation, so I didn't see anything that would flag to me regarding the manner or the completeness with which Oxford replied to the FOI.

Q So if we go to the bottom of page 20, Dr Chaput sets out her concerns about Great Ormond Street Hospital. Now, I think probably it's important to go over the page at this point and just look that she's provided the top of the data set.

Now, if we just notice the columns on the right-hand side before we go back and read her text, there's a column called "Samples", there's a column called "TotalOrg", and column called

"Episode14Day". If we go back to page 20, what I'm going to propose we do is that we look at the cover document for Great Ormond Street, so that's bundle 44, volume 10, page 6, and what I want to do is to go to page 8, which contains the answer to Question 6.

If we could zoom down, so the bottom half page if as visible as it can be. Thank you. We've asked the question, but they've slightly cut down the text, so it now reads:

"A list of the numbers of all organisms, by species, isolated from blood cultures and patients on the paediatric haemato-oncology unit (whether deemed significant or not), by site (peripheral venipuncture, peripheral line.

Please see the table below which shows the numbers of all organisms, by species, isolated from blood cultures from patients on the paediatric haemato-oncology unit (whether deemed significant or not) by site (peripheral venepuncture, peripheral line…"

If we jump over the page, we hit the data set, so if we go back again, please. So if we go back Dr Chaput's document and we look at her criticism, she discusses her view that the first column-- This is the third line from the bottom:

"In short, the first column gives the first total positive samples (no deduplication) while the third has been deduplicated by episode. If Mr Mookerjee had indeed used deduplicated data, he would have focused only on the third column."

So if we go on to the next page and we take the top of the page, please. Just get right up there. Just keep going further, please. Thank you. Now, when you were working out what information to put into your total for Great Ormond Street, which column did you use?

A So I have to say that I did this early on in 2024, so excuse the broadness with which I'm going to, well, answer this question, but I remember sort of taking into account the infections which were attributable to environmental bacteria and fungi and using what I think is a combination of TotalOrg and Episode14Day.

Q So how would that combination work?

A Essentially, it would work in that I extracted the data into a spreadsheet, wrote some code on the software called R, which we are now familiar with, write in assumptions and then ask for an output.

Q What would the code do to these two columns?

A What I wanted to-- or what I asked the code to do was to compare these two columns because I recognised

then that I was not particularly clear about what they meant by Episode14Day and how was that so different from TotalOrg.

So what I asked the code to do was that, effectively for every line where there is blood culture of interest, to take into account both these columns, report any inconsistencies, but to essentially, well, give me a deduplicated list, or give me a list----

Q So how would that work? So let's take an example of Aeromonas, which is on your list environmental organisms but doesn't occur at Schiehallion, so that's 2022, Aeromonas. There are three lines. They are quite a lot on the table, and the first has a source of "Peripheral", the second of "Picc line, single", the second, "Picc line, white". In all case, there's a 1 in the Samples column. Am I right in thinking that each of those involve a single sample?

A Well, I think here is where I did not understand it in those terms at the time. Now, I should say that I have revised my understanding in light of the points raised by Dr Chaput of this table.

Q I think it would be important that we do both your original understanding and your revised one for completeness.

A Sure.

Q So what was your original understanding?

A My original understanding was that there is no reason to see, for example, if we take into account years 2022 and the organism Aeromonas and the fact that they-- one is from-- or two are from Picc lines and one is from peripheral, as indicating that these are three unique samples, if you see what I mean.

Q So you think they are three unique samples, or you think they aren't----

A No, I didn't explicitly state within the code that they were three unique samples.

Q Right.

A So I left that ambiguous.

Q Okay.

A Which, in retrospect, I should not have, now that I have----

Q Because the point that Dr Chaput makes is that you've listed four.

A Yes.

Q The point she makes, which peripherally or superficially seems attractive----

A Yes.

Q -- that if you look at these three rows, you can get to 3, you can get to 5, and, at least mathematically, you can get to 2. How do you get to 4?

A I agree. I think once that was pointed out to me, and other examples, that wasn't the only example, but it's a--

that that the code should have outputted one of those numbers. If it took into account the samples, it should have been 3. If it took into account the total organisms, it should have outputted 5, and if the Episode14Day, then it should have outputted 1. That is----

Q Why wouldn't it output 2? Sorry, I don't understand.

A No, sorry. I was looking at-- I should put my glasses back on actually. Not 2, sorry. 3 or 5 or 2, sorry. Not 1, and this, I guess, is, you know, something that I take on board, that I should have been more explicit with the code that I wrote and that somehow the code has jumbled up and in some cases has spat out a number for an organism which neither fits here nor there.

In recognition of that, when I received the points from Dr Chaput and I read through her criticism, which was actually very helpful, and I should sort of just note for the Inquiry that that sort of back and forth and-- you know, that sort of iteration is actually what we do in epidemiology, well, all the time. You would do something. You would then have colleagues who check it. They would have a few things to say about it. You then improve on your analysis.

Unfortunately, as part of the Inquiry, that takes the form that it does in terms of things being submitted and then things being looked at and then, of course, you have colleagues who then understand the math and understand the data, then come back to you.

So I accept, I recognise and I took into account the rebuttal and in my most recent analysis, I have gone with what was recommended by Dr Chaput which is, in her opinion, that the Episode14Day column for GOSH (Great Ormond Street Hospital) is the one to take.

Q Can you help me with something that I simply don't understand?

A Sure.

Q I understand that that's what you've done. But let us use an example of an organism, and I don't know whether it's in the environmental-- in fact, we'll use the second 2022 Aeromonas. So, the second 22 Aeromonas is a Picc line single, one sample, one organism.

Now, at a superficial level, I can understand that there's one sample in which they find one organism. When it drops to 0 in the final column, is that indicating that that particular row is possibly duplicating another row for Aeromonas and therefore you should drop to 0?

A Yes, so looking at it now on face value, that is what I would take away from this, yes.

Q Okay. So, although it's a bit

weird that you've got a 0 in an Episode14Day column, you're effectively saying that's because this row drops out because it duplicates another row?

A Yes, because of what-- and I go back to the returns from the other comparators, that they were clear about the final number per organism once they had done the deduplication process.

Q Yes.

A Here, GOSH have gone above and beyond and almost, well, provided what I saw as their workings. So, in terms of, you would normally be looking at samples, the total org and, you know, you would have the stats package then give you a column where the package is making the distinction between if something is to be accepted because it is unique or not accepted because it's a repeat.

But what I would have expected from GOSH, so ideally, is in keeping with the other comparators to just give me the final number.

Q Right, before we go to how you took account of Dr Chaput's work, I wanted to go back to the comparator chart that you did last year.

A Sure.

Q Bundle 21, volume 1, page 86. So, if I understand correctly, GOSH is the dark green.

A Yes.

Q In fact, it's the highest rate in each of ’15, ’16, ’19, ’20, ’21 and ’22?

A For most of the period, yes.

Q Yes. So, what would be the effect of deduplicating it in the way Dr Chaput has done on its place in this chart on GOSH? Would it pull it down within the groups?

A Absolutely, because-- well, first of all, I should just say a short statement that when you look at GOSH here with-- you know, with the caveat that we have discussed regarding how I calculated being a black box to begin with. But actually, yes, the GOSH rate is one of the highest. But in being the highest, it's not so different to the others, in that it doesn't stand out being explicitly or exaggeratedly high. So that's one.

Q Right, okay.

A If I-- and we will come to the analysis bit, the chart bit, if you will take into account the, as is suggested by Dr Chaput, 14 dedupe column, that rate for GOSH will go right down to being one of the lowest.

Q Okay.

A So, essentially, well, exaggerating the difference between the Schiehallion rate, whatever that rate might be, and the GOSH’s rate as a comparator.

Q Well, my Lord, this might be a good point to have our morning coffee

break before we turn to this new GOSH rate.

THE CHAIR: Yes. Could I ask you to be back for five to twelve, Mr Mookerjee?

(Short break)

THE CHAIR: Mr Mackintosh.

MR MACKINTOSH: Thank you, my Lord. Now, Mr Mookerjee, just before the coffee break, you indicated that you had reviewed Dr Chaput's critique----

A Yes.

Q -- of the handling of the Great Ormond Street data and that you had a proposed course of action. What was that?

A So the proposed course of action was for me to take into account the data, in Dr Chaput's opinion, which should have been taken into account by me with regards to the GOSH set of data, which is to focus on the column 14-day dedupe or 14-episode dedupe.

Q And is it, in essence, just to simply take that column?

A It is essentially to take that column as representing the infection incidence of environmental bacteria and fungi for GOSH, recalculating the rate of infection per year for GOSH, but----

Q But only for the environmental bacteria and fungi that meet the case definition?

A Absolutely. And then, for completeness, taking into account what I call the new GOSH rate and use the numbers and aggregate it with the other three comparators to come up with a new-- well, overall comparator rate for each of the years from '15 to '22.

Q Okay, so I won't take you to your calculations yet, but just to be clear, what you've effectively done for Great Ormond Street, you've taken Dr Chaput's view, you take the right-hand column only for the case definition organisms, total them up by year and that becomes the GOSH contribution to the overall rate?

A Yes.

Q Right. The admissions stay the same?

A Yes, the admissions remain the same.

Q Now, you produced two reports in the last week.

A Yes.

Q One of them which I'll show to you is in bundle 44, volume 9, and it is at page 3, and this was produced-- is your first response to Dr Chaput.

A Yes.

Q It would be fair to say this one doesn't contain an acceptance that you should use the right-hand column in the GOSH data as Dr Chaput proposes?

A Absolutely. No, it does not

contain.

Q So why the change of position since-- what was it? Tuesday?

A Well, a recognition that the simplest way to understand this and to help the Inquiry in getting to a point where it has the evidence it needs regarding the rates of infection and how that compares to the comparator, and to avoid the back and forth about the methodology and to put the criticisms to rest is to take into account what is being noted by Dr Chaput and the time that she-- that she had put into it in formulating the response, and take for GOSH the 14-day dedupe, and if I may add, go further and--

Because, of course, the-- So taking into account the criticisms of the deduplication that was carried out by me when calculating the Schiehallion incidence and rate using the GGC spreadsheet and to put those criticisms to rest is to take what I note are the incidence figures from Dr Chaput.

Q So if we just stick with the comparators for the moment.

A Sure.

Q You've explained why you've now moved to use her Great Ormond Street numbers, and we'll see the numerical consequence of that later, but if we could turn to the GGC data set, so back to her document, so I think it's page 23 of this bundle-- no, page 22.

Right, so she's created a table, Table 1, which runs over three pages. For each of the infections that meet the case definition, she has provided three columns, the first of which is your count, which comes from Dr Mumford, actually, the second of which is the total number of infections she can find, and then it's her deduplication. Now, you produced a response to this, which is in a similar style on page 7 of the same bundle.

A I did, yes.

Q And you've gone through: agree, disagree and, indeed, the right-hand column is there because I asked you to provide some specification of why you disagreed.

A Absolutely, yes.

Q Would it be possible-- Well, let's understand how you deduplicate the GGC dataset, so take that off the screen for the moment. Now, I can't put on the screen the GGC dataset because what's in the first column?

A CHI numbers, which are unique for each patient.

Q Yes, and is there a column – I now can't remember if it's L or O – which lists the numbers-- the types of infections found in a sample?

A Yes, there is a infection code column which gives you the sample-- which gives you the organism codes, so

essentially the code corresponding to the organism for which the culture was positive.

Q Yes, and whilst it may not, at the end of the day, matter for the numbers, I think it matters for completeness: have you any views of difficulties you've had in interpreting the GGC spreadsheet for the purposes of deduplication?

A Yes, so, I mean, I think I-- even in the quantitative report, which was my first report that I wrote, well, I tried to be open and honest about the difficulties that I had in terms of, first of all, understanding the dataset and then to using it to produce a deduplicated list of environmental infections for the Schiehallion.

And I think in my reports that you had-- just had on the screen, I list two things which caused me the most problems in terms of adjusting for them. One is that if a sample was positive for more than one organism, of which there were many, 1,000----

Q How many rows did the spreadsheet have?

A In total, the spreadsheet that I got had 216,000 rows. I think it was 215,800, so give or take, rows of data, of which 1,768, well, give or take, had in one line multiple organism codes in one cell, so under the column-- so organism code, separated by a space or a comma but in one cell.

So, essentially, these were numerous instances of polymicrobial samples whereby each and every organism for which that sample was positive had its organism code entered not in three separate columns, so designated, as I'm used to, as Organism 1, 2 and 3 from a Sample A, but within one cell.

And, as I said, I was dealing with 215,000 rows; 1,700 of these had at least two organism codes per cell. There were a lot of instances of the polymicrobial-- the positives extending into 3 and 4s, which was humongously, well, complicated to split up because essentially what I would've expected to be given is a row unique to every patient sample and positive.

And, therefore, for a polymicrobial positive sample, which, for example, had three organisms identified, you would have three rows.

Q In this field that had multiple codes, was there always only one code used for each organism?

A No, so that was the second issue, which was that what confronted me was multiple codes for the same organism within the spreadsheet. So essentially, I received a spreadsheet that had not been cleaned or formatted in a

manner that I'm accustomed to, to essentially aid the analysis.

Q How do you respond to the suggestion that Dr Drumright would've had the same sort of table, perhaps not all the same columns, but with these polymicrobial cells, just like you did? How do you respond to that suggestion? She didn't seem to have a problem initially with deduplication, albeit she did when she provided us with later data, but she didn't initially have a problem. How do you respond to that?

A I think the only, well, honest way that I can respond is that in an ideal setting, you would be able to take the aid of a colleague to help in these matters. And if Dr Drumright and myself had done this together, she would've been a big aid in helping me to analyse the data set.

As it was, it was just me looking at 215,000 columns-- sorry, rows, and getting my head around, to the best of my abilities again, the problems that I encountered with the same role for multiple organisms, and in that, further confusing matters, not the same organism code for every instant when that-- when an organism was found.

Q Did you eventually have cause to check your coding?

A Yes, I did.

Q What did you find when you checked your coding?

A I found that the code I had used in R had not done as good a job as providing me with a unique patient sample infection set as I would have wanted.

Q Did that then affect onwards the deduplication stage after that?

A Yes.

Q Right. I wonder if we can go to page 24 of bundle 44, volume 9, which is the bottom of Dr Chaput's table. When she's done her checking, she observed that your total number of infections over these years in the Glasgow sample is 187, and she thinks it's 159.

A Yes.

Q Now, when you reviewed her analysis, what proportion of her criticisms would you accept are-- I mean there are 27 differences there. What proportion of that difference would you accept is validly made by Dr Chaput?

A Well, I think I, in my response and now, agree with the vast majority of the criticisms, and therefore accept the numbers that in the Dedup, well, column in particular that she provided to be the numbers for those environmental organisms. And I should just well underline the word "vast majority" because we----

Q There's some you don't accept, but you accept the vast majority.

A Yeah.

Q Right, and so you explained to us a few moments ago that you decided to accept-- to work with her numbers.

A Yes.

Q Now, if we look at her report, she doesn't do it by years, does she?

A No.

Q This document, which ends on page 24, isn't an annual total for each organism.

A No.

Q If you're going to work out an annual total for Glasgow for each organism, how do you do that using Dr Chaput's table?

A So I think it's important now to look at the table that I produced in response to this table.

Q Do you mean page 7 in this bundle?

A Whereby, when we get to it----

Q This one?

A Yes.

Q Right.

A You would see that my reanalyse numbers under "14-day Dedup", which you would aggregate and would be your final figure.

Q So that's the fourth column from the right, under the heading "Mookerjee latest".

A Yes.

Q Under "14-day Dedup", it goes 1, 1, 4, 3, 5, 1, 1, 1, 1, 1, 6, 1. Right.

A Absolutely, and to be able to do that, what I have had to do is to look at the notes that had been provided by Dr Chaput, well, hopefully, in terms of the total numbers that that arose from her analysis and the ones which were unique, to work out then what makes up her incidence figures----

Q Per year.

A Per year for those organisms.

Q But we don't find that on this table because----

A No, we don't.

Q -- this is the earlier iteration.

A Yes, absolutely, so I have arrived at that because I have her final number of 159. I was pretty, well, close to it at this point in terms of my figures, and she has helpfully, well, provided where we-- where she had not agreed with my original numbers----

Q So that's page 22.

A -- what her numbers were based on. So essentially, I could tell the difference between what I was saying the number should be and what she was saying the number should be.

Q Is that by reference to the "Notes re duplication or counting errors" column in her Table 1 on page 22 of this bundle?

A Absolutely, yeah.

Q Right. Now, I understand you were doing this recalculation yesterday.

A Yes.

Q So last night you provided us with another document.

A Yes.

Q I'm just going to put that on the screen. It's not yet in the bundle my Lord. If we start with the heading page. There's a header page on it. I've given a hard copy, my Lord, to all the Core Participants in the room, and it will make its way into bundle 44, volume 9 as an additional document. There'll be a reissued version of 44, 9 from our document management team later in the week, so people will have to download a fresh version of 44, volume 9.

(To the witness) But the front page is your summary, and at 1.2, do we see that you've noted that in her report Dr Chaput "calculates the total Schiehallion infections as, n= 159 for the period 2015 – 2022"?

A Yes.

Q She's provided a table, and using her table, you say you've calculated a breakdown of 159 as a yearly incidence figure.

A Yes.

Q We've already dealt with GOSH, so we'll pass over 1.4 and GOSH. Now, if we go over the page onto the next page, there's two tables, the Table 1. What is Table 1 trying to show us?

A So Table 1 was intended to show you the incidence, well, figures which had been calculated.

Q These are calculated by you?

A Yes, and these have been calculated by me in recognition of what has been flagged by Dr Chaput, in that the second column is what we have discussed already, that I've taken into account the 14-day dedup column----

Q So that's the right-hand column in the----

A That's the-- Yes, that's----

Q In the GOSH-- Remember, his Lordship's going to write this down. On the right-hand column, in the Great Ormond Street infection data return.

A Yes.

Q Right.

A And those are the incidence figures, which I term the new GOSH infection incidence figures for each of the years, from '15 to '22.

Q And you add that into the others to create the new overall rate?

A Yes, absolutely, and then on the column on the right of that, which reads "Dr. Chaput Schiehallion infection incidence by year", that is my analysis of the breakdown, based on Dr Chaput's total number of 159 and the differences that she made clear between my original numbers and her numbers, the incidence of infection at the Schiehallion for each of the years, from 2015 to 2022, totalling

159.

Q So that column in Table 1 is a replacement, effectively, for the table we see in your original report from last year at bundle 21, volume 1, page 24, paragraph 8.1,15, the right-hand column?

A Yes, you're right.

Q Okay, so we actually have a lower rate in 2015 from Dr Chaput.

A Well, here we're comparing the incidence, so we have a----

Q Well, no, if we go back to the previous page in the new document, that is that an incidence or a count, sorry?

A The incidence, so the count.

Q It's the count.

A So it is-- the infection incidence is a count.

Q So just numbers.

A Just numbers.

Q Just numbers. So if we flip between the two, 3 is what you now have, but before that you had 7.

A Yes.

Q Then before we had 27 and 16, you go back. We now have 25.

A Yes.

Q And then 2017, you had 52 now, but previously you had 66.

A Yes.

Q So it brings the '17 number down.

A Yes.

Q At '18, you had 44. Bring it back. You now have 41.

A 41.

Q '19, you now have 18. Flip it back. Previously, you had 19.

A Yeah, had 19.

Q 2020, you had 9, and now you have 7.

A Mm.

Q '21, you have 7.

A Yes.

Q Now, over the page, please, we have 8. Previously in '22, you had 7, and now we have 6.

A 6, yes.

Q Right. Now, what I wonder if we could do is talk about statistical significance.

A Sure.

Q Have you sought to-- Can you help us about whether there is anything significant between the rates of infection calculated by these for the Schiehallion rates, whether your ones from last year or Dr Chaput's, and the overall comparator rate, whether it's the original overall comparator rate or the new one recalculated by her?

Is there a method by which we can see whether in any year there was a significant difference between the Schiehallion rate and the overall comparator rate?

A Yes, so you do that by taking into account, for the purposes of this

argument, the numerator and the activity for each of the years for the unit in question and you take the numerator and the activity – the aggregated one, I should add – for the comparators----

Q Yes.

A -- and you then do the analysis on R as I did and you do a very, well, basic analysis where you calculate what is called the rate, the difference in the rate between the unit in question and the comparator for each of those years, and that difference is called the rated ratio.

Q The rated ratio?

THE CHAIR: Sorry?

A The rate ratio.

MR MACKINTOSH: The rate ratio?

THE CHAIR: The rate ratio?

A The rate ratio, which you can see in Table 2, well, column 4 from the left.

MR MACKINTOSH: So if we don't worry about what's being compared with what at the moment----

A Yes.

Q -- and just look at how it's displayed. So, the left-hand column is the year, and are the second and third columns the two things you're comparing with each other, effectively?

A Yes. So, yeah, the second column is the rate where we are trying to understand whether it's low or high. The third is what we are comparing that rate to, to understand whether the first was, so high or low. The rate ratio is essentially the difference between the proportions for each of the years in question.

So, for example, for sake of argument here, the rate ratio of 0.26 is an indication of what is the rate ratio, i.e. the difference in the rate, between the unit rate of 2.30 and the overall comparator rate of 8.96.

Q So it's telling you that, in that particular comparison, the Schiehallion rate is a quarter of the comparator rate?

A Yes, absolutely.

Q Right. Now, what does the CL low and CL-- or is it CI high----

A So it should say CI----

Q Right.

A -- and CI-- and CI stands for confidence interval, and the confidence interval----

Q Are these the shadings we've seen on charts produced by you and Dr Drumright?

A Yes, absolutely. Yeah, and the confidence interval is a -- so, when you would take into account the confidence interval low figure and the high figure, it gives you the range of values within which the true estimate should lie.

Q Is this driven by the comparator's number?

A Yes, so the more-- the higher the numbers you are comparing, and by higher I mean just the quantity and we address this whereas for-- at each, well, point in time, we were comparing a large number of infections from the comparator, the aggregated infections and the large number for the activity as aggregated activity for the comparators. You can see that that reflects itself in in the narrowness or the broadness of the confidence interval.

Q So, just to check, confidence interval is saying, if these two are linked then it will fall within those ranges, 95 per cent of the time?

A Yes. I mean, what it is saying is that the rate ratio of 0.26, as an estimate of the difference between the two rates, falls within 0.08 and 0.81, and within those two values of 0.08 and 0.81, we understand where the true value to lie, and therefore, since 0.26 lies between in, that we accept it as a meaningful value.

Q An expected value?

A An expected meaningful, so, value.

Q Then what does the “P-value” column tell us?

A So what the p-value does is that it gives us an estimate of the statistical significance of the difference between the two rates that we compare.

So here what it’s saying is there’s a p-value of 0.02 and we have heard in this Inquiry that p-values of less than 0.05 are understood to point us to an understanding that, in this case, the difference between the two rates is significant, i.e. there is very little - what's the word? The possibility that this difference is due to chance is very, very low.

So, essentially, the possibility that the difference between these two rates that we find is-- due to chance is less than 0.05.

Q Is that related to the idea that the confidence interval represents 95 per cent of the expected locations?

A Yes, so the confidence interval is telling you that this rate-- this estimate of the difference, which is the rate ratio, falls within the confidence interval of 0.08 to 0.81 which is not too broad and that in conjunction with the p-value which gives you further-- which gives you further certainty or, so, it might be uncertainty about that estimate.

So here it gives you the certainty that because the p-value is less than 0.05 that we can be really certain about the fact that there is a real-- that the rate ratio as a quantity is real, so that the Schiehallion rate in 2015 was significantly lower than the overall rate.

Q In this comparison?

A So, in this comparison.

Q Which I recognise might not be the one you want us to use?

A Yeah. Yes.

Q So, just using this as a sort of checking we're understanding tool, so that I can just check I've understood this correctly, you divide the comparator rate, which is the third column----

A Yes.

Q -- by the rate you're interested in, which is the second column, and that gives you 0.26, or is it the other way around?

A No. So, not exactly.

Q No?

A Because you don't compare rates. You compare the proportions.

Q Right, so you compare the proportion between the comparator rate and the thing you're comparing it with?

A Yes.

Q To give you the rate ratio?

A Yes.

Q The CI-low and the CI-high are a range of rate ratios?

A Yes. So, that is the range within which, yes, 95 per cent of the estimate should lie and the rate ratio lies within it.

Q If it was random?

A Yeah, if it was.

Q Right, and then the p-value tells you that the comparator rate of, in this case, 2.3 lies outside that range of rate ratios in a manner that you would describe as significant because it's got a p-value of less than 0.5?

A Well, mostly but I'll----

Q Please clarify.

A So I'll correct that statement or the best way to put that statement is that-- how you would put it in epidemiological terms is that there is a statistical difference between the overall comparator rate and, in this case, the Schiehallion rate for that year, and that that statistical-- the difference is significant because the p-value is less than 0.05.

Q Okay. Now, what I'm proposing to do now is actually to move to the next chart because, if I understand correctly, the comparison you actually want us to look at is on the next table?

A Yes.

Q Right. The next page, please. So, we could zoom this in so it fills the whole of this panel of the screen. That would be helpful. Now, you've entitled this, "Dr. Chaput's Schiehallion incidence rate". Now, to be fair, would it be fairer to describe this as, "A Schiehallion incidence rate inspired by Dr Chaput"? It's not actually her rate by year.

A Yes. It is inspired by the details that she has provided. Well, both in terms of the total figure and the

detailed notes she provided where my numbers had not matched hers by organism.

Q So, just to get her-- I mean I don't know whether she's watching this, but if you've got the same total number of organisms as her, 159?

A Yes.

Q Yes, and you hope you've got the right number in each year that she would be happy with. You don't actually know that, but you've based it on her table to do that.

A Yes.

Q I mean, she could have calculated an annual rate in her document, but she didn't do that. So, you're working out one yourself.

A Yes.

Q Right, and the new overall comparator rate takes account of this new Great Ormond Street incident which uses the right-hand column as she proposes?

A Yes.

Q Right. So, what I want to do here is to understand what this actually tells us, in your mind----

A Sure.

Q -- and then to interrogate it in more detail. So, using the first row as an example, can you walk us through the 2015 row explaining what each column is trying to tell us and what the conclusion you reach at the end of this row is?

A Yes, sure. So, for 2015-- so, I'm going to go from left to right.

Q Yes.

A For 2015, so column 2, we have a rate of infection that is calculated per 1,000 admissions using the information we have on Dr Chaput----

Q Yes.

A -- and that rate comes to 2.30.

Q Per 1,000 occupied bed days.

A Per 1,000----

Q Occupied admissions.

A Yes. It's a correction, yeah. Per 1,000 admissions. So, for that year, then corresponding to that figure, we have the new, well, overall aggregated rate, which, to remind everyone, so, takes into account the absolute values of the new GOSH incidence using the 14-day dedupe column.

Q Just to recap, it also includes your assumption that the others have deduplicated----

A Yes.

Q -- and it includes your treatment of 2016 in Leeds which Dr----

A Yes.

Q -- Chaput's criticised? That remains unchanged?

A Yeah, it remains unchanged in that I do not account for-- as in my response to the criticism with regards to the Leeds 2015 rate, it does not account

for the admissions----

Q I think it might be 16.

A So, sorry, yeah.

Q It doesn't account for?

A No, for the admissions. So, the issue has been corrected.

Q So, for 2016, it will pull that number down, but not for any other year?

A Yes, absolutely.

Q Right. So again, you've dealt with 5.53. What is rate ratio 0.42?

A So, the rate ratio of 0.42, as we discussed, is calculated by, or is a value of the difference between the new overall comparator rate and the Schiehallion infection rate for that year. So, essentially, you're comparing 2.30 and 5.53, but, as we noted, you don't run the rates, you run the proportion. So it's the numerator and the activity that makes up 2.30 being compared to the numerator and the activity that makes up 5.53.

Q So you're comparing four figures, not two?

A You're comparing four figures. Absolutely. You're not comparing two. Absolutely. Moving right, CI-low is the low, well, confidence interval, so, value.

So it's the lower point in that range and the confidence interval high or CI-high is the high value within that confidence interval of 1.33. The rate ratio of 0.42 falls in between 0.13 and 1.33. So, essentially, as we have seen in those charts, it's that shaded figure around the estimate.

Q Yes, and that gives a p-value of?

A That gives you a p-value because we ask whether the difference in rates is significant from the p-value point of view, it gives you a p-value that is higher than 0.05 and 0.14. So, what we say from that is that the difference in the rate is not significant.

Q So, it may be low but it's not significant.

A It may be low, and we can see from the rate ratio that it is about half. So, what you can get from the rate ratio, in layman's terms, is that the Schiehallion rate was about half of the new overall comparator rate, but that difference is not significant.

Q Okay. So if we look at the next row-- Well, I wonder if we could look at-- whether you want-- No, look at 2016. Talk us through this one. This is the one that's got the Leeds partial year in it. So what have you done in 2016 and why is it green?

A So 2016 is green because I wanted to highlight any of the-- or any and all of the columns where the p-value was less than 0.05 and therefore noted that the rate ratios being calculated were significant. Here, in 2016, as per my calculations of Dr Chaput's Schiehallion

infection rate, the rate is 11.03.

So, going right, the new comparator overall rate, so taking into account the new, well GOSH rate, is 4.49. The resulting rate ratio, which is the difference between the two proportions, is 2.46. The confidence interval is between 1.57 and 3.86 and the p-value to two decimal places is 0.00----

Q Sorry. I don't understand, Mr Mookerjee, why a rate ratio of 2.46 which sits between 1.57 and 3.86 is not slap-bang in the middle of the confidence interval.

A Because the confidence interval is an estimate of where-- I mean, it essentially is trying to say is that 95 per cent of the time, if you rerun this, you know, so all those times that rate ratio should fall in.

It's not an average or the rate ratio of 2.46 is not the middle value between CI-low and CI-high. CI-low and CI-high are giving you the range of estimates and those two values will be closer together the more the data that the rates are based on.

So, essentially-- how it goes is, if you have a lot of data, the confidence interval around the estimate will be really narrow, or narrower,
as compared to if you have less data where the understanding is that, because you have less data, the estimate has high confidence intervals around it saying that, actually, there is more variability in what that estimate truly is.

Q Well, I get that. The thing that's confusing me, and it may just be me being not very good at this, is that if we look-- I think it applies, actually, to every single row on the chart. The rate ratio sits somewhere between the CI low and the CI high value in every single occasion.

A Yes.

Q Now, why does that just not mean that the rate ratio is within 95 per cent confidence or there's nothing statistically significant happening here?

A Because the manner in which you-- Because confidence intervals are giving you an understanding of how good your estimate is.

Q I think I misunderstood, but look at the 2016 row. The rate ratio between 11-- I know it's not between the two rates, but for the purposes of just seeing it, the rate ratio between the Schiehallion infection rate calculated by numerator and denominator coming to 11.03 and a comparator range of 4.49 is 2.46 according to your calculation. Does that mean that 11.03 is 2.46 times 4.49?

A Not exactly.

Q I see. Right.

A Not exactly, no. If----

Q Because that's not how it

works?

A No. No, it's not how it works. The rate ratio is-- Sometimes if you-- if people don't go ahead and calculate from the rate ratio the confidence intervals of that estimate, and then the p-value will give you the significance of the difference.

Q So what we're not doing is simply looking at the rate ratio and checking it out with the range of CI high to CI low?

A No, we're not doing that. The----

Q Right, however much attractive we might have thought that, the way you----

A Yes. No. No, absolutely not, and confidence intervals are tricky things to get your head around, but confidence intervals are giving you a feeling of how good your estimate is or how much value you can put on your estimate.

Q Will that be determined to some degree by the size of the two comparator numbers?

A Yes, so higher those numbers-- So, for example, for-- just for argument's sake, if we were comparing 1 divided by 2 to 3 divided by 4----

Q Can I give you a slightly more accessible number for our purposes?

A Sure, sure.

Q Let's imagine that there was a comparator which had 100,000 admissions and 10,000 infections----

A Yes.

Q -- and you were comparing it with something that had a similar size number of admissions and infections, so say 5,000 infections and 50,000 admissions.

A Yes.

Q Would that produce a different CI low and CI high if the comparator was still 10,000 and 100,000 but the thing you're looking at is actually 1,000 and 5,000, a much bigger difference in size? Is it the difference in size between the comparator and the thing you're comparing that's driving this?

A Yes, it's the size of each of those proportions which will drive the confidence you have around your estimate, which is what the confidence interval is telling you.

Q So, in this case, the bigger the number of admissions in the comparator and the bigger the number of infections in the comparator compared to the number of admissions and the number of infections in the Schiehallion sample, the narrower your confidence intervals get?

A And therefore the more weight you can give the estimate.

Q So we shouldn't look at the rate ratio and just see, "Oh, it fits in between…" That's not how it works?

A No, it's not how it works. It's giving you an estimate. It's giving you a-- it's telling you how much can you trust the rate ratio estimate.

Q So, again, I may be going down the wrong end of a long rabbit hole. The units of the CI low and the CI high, are they the same units as the comparator rate?

A No they're not.

Q Okay.

A Because the unit in the CI low and CI high are the same as the unit for the rate ratio.

Q But then why don't you just drop the rate ratio in the middle and say, "It's in the range, I'm fine, I'm going home now, there's nothing to see here"?

A Well, that is essentially what the confidence interval is telling you. It's telling you that 2.46 lies between 1.57 and 3.86.

Q But then it wouldn't be statistically significant because it lies in the middle of that space.

A No. When you talk about statistical significance with regards to the p-value, it's slightly-- so different when we are talking about the confidence intervals. You are measuring two different things, or those numbers are telling you different things.

Q So, in this case, you've got p-values for 2016, 2017 and 2018 that are 0.

A Yes. Well, if I had extended the decimal places you would get something like 0.003 or 00. I chose to keep everything to two decimal places. But yes, the p-value here is saying that whatever you have given in terms of calculating the difference is sufficient in terms of the numbers for the p-value to indicate with certainty that there is a difference.

Q In those three years?

A In those three years.

Q Now, I recognise you did this yesterday at short notice, but given the rigour that we want to apply to this, might it have been better to produce this table with four extra columns – those being the number of infections, the number of emissions, the number of infections – for both datasets? Because then you'll be able to see the whole-- then someone else will be able to replicate your work.

A Yes, yes, well, absolutely, and I was happy to make that available.

Q What I'd like you to do, because I'm conscious that a number of people will be thinking, "Let's check this"-----

A Yes.

Q Now, for better or worse, this whole outing has happened now rather than last year, but looking where we go, I'm keen that those Core Participants who

have the means to check this can check it.

A Okay.

Q It occurs to me that I may have ways of doing this myself, but effectively am I right in understanding that if we look at 2016 as a row, if we had four extra pieces of data, that is the number of admissions in the Schiehallion Unit that year-- We actually have that on the previous page.

Sorry, the number of infections which I mean we have on the previous page because we know, if we're going back one page, that in 2016 it was 25.

A Yes.

Q So we have that number.

A We have that number.

Q What we don't have is the admissions total easily accessible here.

A Not here, but they are the same admissions as -- what was-- I think, well, put into my supplementary report.

Q Yes, but they're not here.

A They're not here.

Q Then if we go back to the next page, for the comparators we have the admissions somewhere in your supplementary report.

A Yes.

Q But we don't actually have a total of the total comparator infection rate do we?

A So not here.

Q No. If we had those four numbers, would any trained epidemiologist be able to double-check your calculations?

A Yes.

Q Right. What I'd like you to do over lunch is to take this Table 3 and add in those four additional columns. We'll go through it this afternoon and that will enable any Core Participant who is concerned about your calculations to check them and to report back to the Inquiry. I may take other steps as well, but I want to make sure if you're saying that the rate of infection in the Schiehallion Unit in '16, '17, '18 is statistically significant that I can check that.

A Sure.

Q So when you do that over lunch break, email that page only into the Inquiry team. We will produce it as a fourth page to this document and we'll talk about it after lunch.

A Sure.

Q But assuming that nothing changes with your calculations, is it effectively your evidence, inspired by Dr Chaput's infection rate and comparing it with the overall comparator rate, taking account of the Great Ormond Street rate, that there is a statistical significance in those three years?

A Yes.

Q Now, should we be in any way nervous about reaching that conclusion for 2016, given the Leeds partial year criticism that Dr Chaput makes?

A Well, I think there are always-- well, caveats to this, and what I would say is, you know, to avoid any sort of anxiety around the rates, that we can focus on the 2017 and the 2018 years, but the 2016 years were adjusted, as was noted, so----

Q Did you watch Ms Cairns's evidence on, I think, Wednesday?

A Yes, I did.

Q Do you recollect her discussing-- I'll show you the chart she was talking about, 27, volume 18-- That one, page 3, yes. So we discussed this with her in evidence----

A Yes.

Q -- and I have it noted that she thought the important thing that could be useful from your exercise was the trend in the overall Schiehallion rate and the trend in the overall comparator rate and she felt that was helpful.

A Yes.

Q She didn't, I think, feel the numbers themselves were important or useful. How do you respond to what she said? Did you watch the evidence? How do you respond to that?

A So I agree. I concur because, I mean, what this Inquiry has had to deal with is a lot of the differences in the definitions, a lot of sort of different admission figures for the activity and, in keeping with why in my initial reports from last year my focus was on the trends, was that-- What we can see from here is that the overall comparator institution rate--

So, essentially, if we had, well, eyes on the ground live from 2015 to 2022 and we could, by some magic, sort of be looking at these four hospitals, we'd be seeing the fact that their rates of infection are-- their overall rates of infection are stable, with a very slight hike, if you could even, well, call it that, so visually from 2016 to 2017, but--

So, overall, for the entire period, the dotted brown line is really straight and now that we have established that, you can-- as we have done here, we'll overlay the overall Schiehallion rate and -- First of all, you can see that that rate is not stable, that it starts off----

Q Well, let's go back and look at your table. So that's the new document, page 3. It's not visual, I know, but these are the numbers you want us to use.

A Sure. So it starts off if we-- when we dig in to count Dr Chaput's inspired Schiehallion infection rate at below the comparator rate, so at 2.3 per 1,000 admissions for environmental infections, where it is clear that the

comparators were having a lot more infections as compared to the Schiehallion----

Q But it's not statistically significant.

A But it's not-- Yeah, not high enough to be different with a significance value.

Q Right.

A But then we see this change in that the Schiehallion rate then goes up to 11.03, but for that year, the overall-- the comparator rate remains fairly stable at 4.49. Go one step further for 2017, the Schiehallion rate, well, goes up again to 20.25.

Now, it doesn't mean that, you know, everything is well with the overall-- the comparator rate because one could say that they were overall at-- the comparator rate also jumped from 4.49 to 8.10. So they are also having infections owing to environmental bacteria; it's not unique to the Schiehallion.

What is unique is or what we can take from that is the difference, so while they jumped from 4.49 to 8.10, Schiehallion jumped from 11.03 to 20.25 and therefore the ratio of the differences change, and then the Schiehallion goes down in 2018 to 16.29. The overall comparators, for their reasons, see a dip as well to 7.92.

Q Much smaller dip, in some cases.

A Much smaller dip, but a dip nonetheless, and then the Schiehallion rate then continues to drop quite swiftly, 16.29 to 7.64, 4.57, and then it stabilises somewhat----

Q And, in fact, the Schiehallion rate ends up below the comparator unit again.

A Absolutely, so it's a representation even when we take into account or there is-- to use that word there, there is a high level of concordance in the trend that these numbers would point to, which is in concordance with the chart that we were just at, whereby the rates at the Schiehallion started below that of the overall comparators, went higher than that of the comparators for '16, '17, '18, thereafter going back to being similar to that of the comparators, and then dipping below it.

So it is interesting and actually quite affirming that Dr Chaput's inspired Schiehallion infection rate-led analysis has such high concordance with my overall Schiehallion rate and the trend in that rate and how it compares to the trend of the overall comparators.

Q Thank you. My Lord, I think this might be quite a good point to start for lunch.

THE CHAIR: Yes. Now, we

normally would take an hour and therefore I would say we would try to be back at five past two, but do you require more than an hour for what Mr Mackintosh has asked you to do?

A No, I don't think I do----

THE CHAIR: And to get some lunch?

A -- Your Lordship. No, I think that should be fine.

THE CHAIR: Right, okay. Well, we'll try and sit again at five past two.

(Adjourned for a short time)

THE CHAIR: Good afternoon, Mr Mookerjee.

A Afternoon.

MR MACKINTOSH: So, Mr----

THE CHAIR: Mr Mackintosh.

MR MACKINTOSH: Thank you, my Lord. Mr Mookerjee, you provided us with a single table.

A Yes, but----

Q Yes?

A I can just see from this printout that one of the columns----

Q Has lost two significant differences?

A No, one of the columns, the new overall-- the comparator rate has lost its initial-- the digit and only has the decimal place, or the digit after the decimal place. It should really say 5.5----

Q Let's go through it slowly so we can do it.

A Sure, sure.

Q So what I'm going to ask you to do is-- I think what happened is you sent us a screen grab over the lunch break, an image rather than a spreadsheet, would you say?

A No, I sent the spreadsheet, but I think it's just-- whatever it did, it didn't fit it all in.

Q Right, well, let's work it through and then work out what we're going to do with it. So obviously left-hand column is the year.

A Yes.

Q First column is the Dr Chaput-inspired Schiehallion infection incident rate, which we find already in Table 1 of your document you produced last night.

A Yes.

Q Then we have the GGC Schiehallion unit admissions, which you've been using since last summer.

A Yes, the second----

Q The second iteration.

A The second iteration, correct.

Q We then have the "Dr. Chaput Schiehallion infection rate", so that's the first column divided by the second column times 1,000?

A Yeah, yes.

Q So, in 2015, there are 2.3 cases per 1,000 admissions.

A Yes.

Q Yes.

A Yes.

Q Third column, we have the "Comparator environmental infection incidence". This is the one where you've taken on board Dr Chaput's comments about Great Ormond Street on the right-hand column, but not the other comments she's made because you maintain the others have all been deduplicated.

A Yes.

Q Yes, and the comparator admissions is the same comparator admissions you've been using for the whole period?

A The comparator admissions, just note that for 2016, as I noted right before lunch, I took into account the-- some time back the flag by Dr Chaput that I incorrectly on one the documents that she was referring to-- I had included the entire year for 2016 for Leeds. Note that I hear that is the corrected version.

Q So that contains four months admissions divided by a third of a year----

A Six months admissions, because Leeds had it from July to December.

Q But it's a partial year anyway.

A It's a partial year.

Q Then you've got what claims to be the "New overall comparator rate", except if we go back to the document you produced this morning, bigger, Table 3, we have it expressed slightly differently as 5.53.

A Ah, yes, these are two decimal places, and then I tried to fit it in by just putting in one decimal place.

Q So if we go back to the document, you've got rid of a decimal place.

A Yes, because otherwise it wasn't fitting in.

Q Right, but it does exist there in the spreadsheet that underlies this.

A Yes.

Q Yes.

A Yes, and just to note that these rates aren't just a calculation of, or the resulting of, 58, you know, divided by 10,490. It is because you take into account each of the comparators, so if GOSH had had 26, it would take that and then add that, add that, add that----

Q No, so it takes account of all four comparators.

A Yes.

Q Right.

A Yeah.

Q But the only difference between this document and the one we had this morning is this one only has one decimal place to the right of the full stop----

A You're right, you're right.

Q -- or the decimal point.

THE CHAIR: Yes, I mean, just following this, I mean, for example, for 2015, what is represented on the new sheet as 0.5 should be 5.5?

A Yes.

MR MACKINTOSH: Yes, on the screen it's 5.5.

A Yeah, on the screen it's 5.5.

THE CHAIR: On the----

MR MACKINTOSH: So, my Lord, the paper copy has been printed out by the Inquiry team----

THE CHAIR: Ah, right. Right, okay. My----

MR MACKINTOSH: -- and we have compressed that row.

THE CHAIR: My error. I thought we were looking at the previous, but I now see that we're looking at the new document----

MR MACKINTOSH: (To the witness) So, just to be clear, to add more confusion----

THE CHAIR: My error.

MR MACKINTOSH: -- what's on the screen is what you gave us.

A Yes.

Q And what's on the paper is also what you gave us.

A Yes.

Q But when we printed out what's on the paper, we lost two decimal points out of this column.

A Yes.

Q When the screen was constructed, we lost one decimal point.

A Yes.

Q But in the spreadsheet, all the decimal points are there?

A Yes.

Q Right, and then the "Rate ratio", "CI_low", "CI_high", "p_value" and "Summary where p value significant" have not changed?

A No.

Q No, right. My Lord, what I'm proposing to do with this is I've taken Mr Mookerjee's evidence of what this means this morning.

THE CHAIR: Mm-hmm.

MR MACKINTOSH: I'm not proposing to go with that again now, except one question of clarification about p-values.

THE CHAIR: Mm-hmm.

MR MACKINTOSH: Then I'm going to invite all the core participants who wish t to look at this spreadsheet, and then if they wish to contact the Inquiry solicitor with any observations that they think would assist us, to do so ideally within the next two weeks. The reason I say that is because there is only, I think, one witness still to come after the end of this week – I hope I'm not disparaging someone I've forgotten – who has the professional background to interpret this.

THE CHAIR: Mm-hmm.

MR MACKINTOSH: I think that witness is giving evidence in the second week of Part 3, and so if anybody in a sense wants to come back and say, "Mr Mookerjee's made some terrible mistake," they should do so within the next two weeks because that will enable us to find out how to present that to you, do that, put it in front of that witness, and then see where that takes us, if anywhere.

THE CHAIR: Yes, I understand that.

MR MACKINTOSH: (To the witness) The only question I want to ask is - take that off the screen and just go back to the document from this morning, Table 3 - I'm trying to understand what the best question to ask about p-values and (inaudible 14:22:18) is.

Now, I'm going to give you a concept that I think might be relevant. To what extent are you trying to find out if there is the existence or absence of a relationship between the Schiehallion infection rate and the overall comparator infection rate by this process?

A Yes, so I think that that is very well put, that that is exactly what we are trying to ask, and to suffix that point with, if there is a difference, is that difference-- will that difference be significant?

Q Okay, and the final observation just to make is that, when you carry out any of the calculations in this document or indeed the one you supplied at lunchtime, it only operates within one year. It doesn't look at the relationship between one year and the next year.

A Yeah, exactly. It will only look at the year, so the p-value, the rate ratio is all for that year.

Q Okay. What I want to do now is just put to you a few of the comments that Dr Chaput has made that we haven't already addressed. How would you respond to the suggestion that it shouldn't actually be necessary for Dr Chaput to have to carry out this check? It should've been obvious in the form you originally supplied it to the Inquiry.

A I'm not sure I understand the----

Q So Dr Chaput, effectively, has had access to the FOI data----

A Yes.

Q -- the original bloodstream infection spreadsheet, all 215,000 rows, your reports or your supplementary reports, Dr Mumford's reports, and she has worked this out herself. How would you respond to the suggestion that it shouldn't be necessary for her to work it all out? It should have been something that was immediately apparent and you should've disclosed it all in your original report.

A Well, I think, first and foremost,

I have gone on face value in terms of what has been provided by Dr Chaput. To my recollection, I have not seen the methodology that she followed in terms of how she arrived at the final number of Schiehallion infections for 159.

Q Other than the comment you relied on.

A So other than the comments that I rely on per organism, I can't recall seeing her analysis of the Freedom of Information incidence and activity and what kind of rate that would provide, so what I have sought to do or what I have tried to do is to aid the Inquiry and to restrict the numbers of back and forth that we have on this by purely utilising her numbers or the approximations or, you know, the ones I've been able to work out from what she has provided and gone by what she has noted should have been the column that I should have taken for the GOSH data.

What I will say is that if feedback was given sooner, then these corrections would have been made sooner.

Q I mean, I think the point she would make, Mr Mookerjee, is that she couldn't provide the feedback sooner. By sooner, I mean last year. I don't think there's any value in working out when this summer she could have provided the response because you'd already given evidence.

But I think the point she would make in respect of last year is that she couldn't provide this response last year because she hadn't realised the way that you and Dr Mumford had defined the cohort. Therefore, the chain of her reasoning starts with her question of Dr Mumford----

A Correct.

Q -- about what's in the cohort of infections and then, from that reason, Dr Mumford's response causes her to check your numbers----

A Correct.

Q -- and then the whole thing unravels. Now, how do you respond to the suggestion that what you should have done is contained extracts from the FOI response data as much more detail in your original report, so that if this issue-- this issue could then have been spotted last year? How do you respond to that?

A So I take that on. I understand the criticism and I accept that in retrospect, and I think what we have come to see in this inquiry that more data is better.

You know, in the field of epidemiology, as an epidemiologist, I am very used to being asked for the final analysis. It's not very often that the nitty gritty about the back end is requested. So I guess, you know, I have learned or have been reminded or I think we have all evolved in the kind of data we want to

look at, if that----

Q When you say, “We have all,” who do you mean by, “We have all”?

A As in, I look at the evolution of the workings from the HAD report, whereby, you know, one of the many-- the criticisms of the original report was that the workings were not provided, that the admissions, for example, or the bed days, the tables that the rates were based on or the graphs were based on were not provided explicitly, and the criticism of that paper was swift, and then there’s been a change in the manner in which the HAD authors-- have sought to be overly, so, transparent about, you know, the back end data, so that everyone can work it out, if you see what I mean.

Q That you see a process that’s applied, not just to you, of bringing the----

A Yes.

Q -- back end data forward?

A Yes, and if that was what was expected of me from the very beginning then, in retrospect, I would have done that. But I think you know, as with everyone else who has gone through their own process of analyses, it’s a process whereby, you know, you produce something. You could ask certain questions. You produce more to clarify. You accept some criticisms, and you make some rebuttals.

Q So the next point I want to put to you relates to the calculation of the number of infections in the comparator units. Now, I recognise that you have, to some extent-- well, you have entirely accepted Dr Chaput’s critique of your use of the GOSH data. But, absent the 2016 partial year issue, you haven’t accepted her criticisms in respect of Leeds, Cardiff and the Vale or Oxford.

I noted that she explained in her evidence – around about column 186, for my colleagues – that because of the concerns around deduplication that she identified, the effect is that the comparison being made between GGC’s data, even properly deduplicated, and this comparator is simply invalid and should not be attempted and cannot produce a conclusion. How do you respond to that?

A Well, I think there’s a lot of assumptions in that statement. So, in that, it assumes that Leeds, Oxford and Cardiff and Vale have not deduplicated their data. Now, I don’t agree with that.

So, to my mind, and I go back to the graph where we were looking at the trends for each of these hospitals and the rate of infection, that what we at the end of the day need to be looking at are the trends in infection and the differences in trends and whether that that clinically makes sense because the epidemiology

can't exist in a vacuum to what we understand, not hypothetically but to have actually have happened.

There were concerns regarding unusual peaks in infections and the epidemiology, in terms of the trends in infections, regardless of the different ways in which the different authors – the HPS authors, by Dr Kennedy, by myself – in a few different ways, from a few different lenses, from a few different sides have gone about to look at the data. I would remind everyone that we've used admissions, we've used bed days, we've looked at all of Schiehallion, my numbers.

We've even now looked at the numbers inspired by Dr Chaput and, of course, also-- the Inquiry heard last week from two of the authors from HAD. In particular, I mean-- and we will come to the GAM analysis - they all show the same thing.

Now, this is one of those very rare situations in epidemiology where we have curated the rate of infection for 2015 to 2022 in every single imaginable way possible. We have changed the definitions. We have had a change of definitions.

We have had a change of numbers, in terms of one larger than the other, one encapsulating the other, one broader than the other. We have even changed the activity, and they all seem to show the same thing. The caveat is they show the same thing in different degrees, but they all show the same thing.

Q How do you respond to the suggestion that because your opinions and the numbers you've used to support those opinions have changed and evolved over the past 18 months in the way they have, that you simply can't be relied upon because you're inaccurate?

A Well, I think I would respond by saying that inaccuracy is a facet of analysis that you will find very hard to exclude, especially when you're dealing with such big sets of data, which often has been paired with communication one way.

So I have been given data without the recourse to having the data presented to me in a way that would be most agreeable to this sort of analysis. That we've been dealing with a lot of data and, actually, it's in keeping with the epidemiological process, that you--

I mean, in epidemiology circles we don't call them-- we don't think of inaccuracies, you know, as anything more than, "We need to identify them," and then the question is, has it made a humongous difference to the output?

If you chart the-- or if you want to quantify the effect of the inaccuracies, in my analysis, one way to do that is to compare all of the charts that I have

produced throughout the process and to look at the trend, to visually just look at the trend, and what it says is that the trend remains so consistent in the infections at the Schiehallion, and that regardless of what you say about the Freedom of Information requests and how the hospitals responded, it is the best we could do.

Well, could we-- could I have been more perfect about it? Yes. Could we have better data? Sure. I mean, we always sort of-- we aim to do so. But this is what we had and I can't see, well, anyone else who has gone through the trouble of getting FOIs responded to from four large hospitals and, to remind everyone, we sent it to about 11 or 12 hospitals to get enough to be able to do this analysis. So not saying it's perfect, but it is more than I think most people have attempted.

Q Okay, thank you. What I want to do now is to move to the work you did at my request with Dr Drumright----

A Yes.

Q -- about three weeks ago. But we'll start with the HAD report. So I wonder if we can go to bundle 44, volume 1, document 1, page 64. So this is at section 7.2:

"Is there an increased level of infections consistent with there being "widespread contamination" of the water system?"

Now, the reason I put that on the screen is I want just to ask you to explain to me what you understand the authors of the HAD report -- what you think they were trying to do? We'll start with adult patients in the BSI sections and then we'll look at the paediatrics. So, in a sense, what's your understanding of what they were trying to do?

A I guess the way that I see the HAD report in the way that it was put together, and therefore will extrapolate, so, backwards from that to understand what the HAD authors were trying to do, is they went about looking at the rates of environmental/non-environmental infections within-- from the perspective of the consultant at Yorkhill and the RHC for adults and for children.

Q So in what way does that differ from the approach that you took for the children?

A Because the question that guided me was, we understand that there are or there is concern regarding the unusual rates of infection within the Schiehallion, what does the data tell you?

So the methodology or the hypothesis that I went on to answer was -- what does the data, well, tell me about the rates of infection at the Schiehallion unit, and I had to then define what I mean by the Schiehallion unit, and alongside Dr

Mumford and with Ms Dempster, we defined it as the physical spaces of 2A, 2B, 4B and 6A.

Therefore, to understand what the rate of infection in the Schiehallion unit is, we would do that by understanding what is the rate of infection in 2A, 2B, 4B and 6A for the period, so 2015-2022.

Q If you look at the HAD report analysis for paediatrics, is that what they did or they did a different approach?

A The approach they took was different because what they ended up with was a rate of infection -- I wrote this down because it doesn't sort of flow off the tongue, does it? They went about, well, calculating a rate of infection in the paediatric haematology oncology patient population aligned to a consultant.

Q Another witness has described it as the rate for the patients of the haematology and oncology consultants. Is that roughly what you're saying?

A I would agree with that.

Q Do you understand the same principles being applied to the adult patients for the haematology----

A Yes.

Q -- consultants there? Right.

A Both to the numerator and the activity.

Q Given the difficulties that you had with identifying appropriate activity measures, what difficulties arise from an approach which looks for the rates amongst the patients of certain consultants rather than in particular wards?

A The simplest way to put it is the inconsistencies or the inaccuracies, in terms of where these patients are and who these patients are in terms of their specialty or-- so, in that-- I asked the question, are these consultant infection rates, are they indicative of the Schiehallion paediatric haematology and oncology infection rate? I would say no, they're not.

Q A moment ago, you sought to use the GAM models of the HAD rates as one of the lists of things that confirm the same trends as your analysis and others. You've listed them.

Does any difference between what they've tried to do in HAD and what, say, you or Dr Kennedy or HPC have tried to do prevent them being used in the same sort of collective comparison exercise?

A No, I think within the broader context of looking at rates of infection-- So, on the proviso that we understand that the definitions are different, that I think for example that my Schiehallion infection analysis, well, probably sits as subset within Dr Drumright's, well, paediatric haemato-oncology consultant analysis.

So, keeping that in mind, I can see

some value in it being, well, part of the analysis that we take into account alongside the HPS reports and those of others, but as long as we remember the caveats and they're not an absolute like-for-like comparison, but the Venn diagram does extend over. You know, the Venn diagram for the HAD analysis will have considerable overlap with my analysis.

Q Okay.

THE CHAIR: Sorry, did you just say Venn diagram?

A Yes.

THE CHAIR: I'm afraid it's some time since I did my Higher Maths.

A Yes.

THE CHAIR: Remind me what a Venn diagram is.

A So the Venn diagram is essentially-- well, best put-- best sort of seen as circles of overlap----

THE CHAIR: Yes.

A -- and so more the overlap-- Well, with the overlap.

THE CHAIR: Yes. You've reminded me.

MR MACKINTOSH: No, I think the one thing I wanted just to get you to talk about in this context-- because I don't think you've actually given evidence about it and I think you're the primary user of it, so we can go-- I'm going to double-check I'm on the right place. Yes, bundle 44, volume 6, document 13, page 201. So this should be the second-- it was actually the first GGC response data.

Now, something terrible has happened to our screens because they're flickering. They've stopped flickering. I wonder if we can simply go, for actual ease of conversation, to page-- Stay on this page, 201, and we'll look at the bottom half of that table, which is January 2018 to December 2018.

When you came to find your geographical connection to work out what was the Schiehallion unit, were there any difficulties in this data that you found, given that this is supposed to be haematology and oncology patients?

A Well, yes, that you have unit admissions here that are not specific to the paediatric haematology oncology cohort.

Q I mean, does it matter? The reason I put this to you is if you look on the right-hand side, if we just look at 2018, because we know there's lots of movement in 2018, bottom right-hand corner of the whole spreadsheet, there is 1,186 as a total number of admissions who stayed overnight.

A Yes.

Q That's the first set you've got, and if we just look along that bottom row, we see some big numbers: RHC Ward 2B, 575; RHC Ward 2A, 178. So they're

what you'd expect, they are the Schiehallion Unit?

A Yes. We understood the Schiehallion Unit to be the physical spaces, 2A, 2B.

Q Then if we go to the far left-hand side, column 4, we have QEH 6ACH inpatients and QEH 6ADCH day unit. Again, that's what we understand to be after the decant.

A Yes.

Q So what do you make of this? Firstly, what do you make of them, and secondly, do they matter? What do you make of the 23 cases in RHC area 1B, day surgery? That's the sixth column, and secondly the 138 cases in the Clinical Decision Unit, and then there's a couple of others smattered around the hospital. What did you make of this when you saw this dataset? What do you think you ought to do?

A Well, I mean-- So I know the answer in terms of what I did, but to rewind back to the impression that the spreadsheet gave me, I mean, one of the first questions was that are these wards and the admissions aligned to them in line with what the title of the spreadsheet says? Is that--

This is the RHC haemato-oncology, from which I understand that these should be admissions that are all specific to the paediatric haemato-oncology, well,

patients. So, in conversations with people who understood about the movement of patients and about how each of these ward areas were used, it dawned on me that there are a lot of ward areas here that are not specific to the paediatric haemato-oncology-- so patient cohort, and so one had to go with what was evident-- so in terms of what was explicitly stated as a ward like RHC Ward 2A, and only take the admissions within that column.

Now, saying that, the other issue that you would see being repeated within the spreadsheet is that the same ward name occurs in more than one column, and that was a separate set of back and forths with colleagues, with Dr Mumford and with Counsel (to the Inquiry) to understand, first of all, why have I been presented with a spreadsheet with these wards when the explicit ask is for haemato-oncology-- well, paediatric patient admissions?

Q Could it simply be that you, and potentially others, have had a slightly simplistic understanding of patient placement in the RHC when, actually, it's just more complicated than you imagined?

A Yes, having worked within, you know, the NHS for quite a few years now, I understand, well-- so completely, and concur, that patients are a lot more

haphazard than we think, but I was not presented with the patient pathways, in that I couldn't look at where each paediatric haemato-oncology patient was for the duration of their stay.

Q Okay. Before we move onto the (inaudible 14:54:49), one more thing I want to show you. We had some evidence from Ms Cairns about occupied bed day data provided by to GGC by NSS in 2019, and occupied bed day data provided to the HAD authors. I'm not going to revisit that with you, but there was a discussion that it might be available in the Public Health Scotland website, and I think you made an attempt to access this information.

A Yes.

Q So if I can show you bundle 21, volume 1, document 1, page 60, and this appears to be you're describing how you're trying to get bed day data from the online system.

A Yes.

Q Firstly, had you used this public health system before?

A No, I hadn't used it prior to me doing that piece of work.

Q Ultimately, the point you, I think, made in this document, that it wasn't possible to pull the data that you wanted----

A No. I think what I say here, and I just have to remind myself of it, is that the Public Health Scotland website does provide you with admission data, with the data for bed days activity, but it only does that with within the subset of certain-- the specialty codes.

Q Right.

A But you can access it.

Q Okay, but this is the website that you tried to access it from?

A Yes.

Q Okay. Take that off the screen, please. I want to go back to the HAD report, 44, volume 1, page 115, and effectively set up what I asked you to do with Dr Drumright by just providing some context. So, if we can go to Figure 21, which is on page 117, and this is described as:

"Bloodstream infections attributable to organisms with no environmental relevance at Yorkhill and Queen Elizabeth with a fitted line showing change over time."

So when you saw these fitted lines, what did you understand these to be?

A I understood these fitted lines to be the trend lines that are fitted to the data points, the yellow-- the dotted one being the trend line which was fitted for the entire set of data from 2005, Month 1, to 2015, Month 1, and then the green, the dotted line, is the linear line which has been fitted for the period from 2016 to 2022.

Q Now, there's similar trend lines over the page, 118, in Figure 22. Did the authors at this stage provide the reader with any information about the level of significance or the quality of the fit between the trend and the line?

A No.

Q But Dr Drumright ultimately did provide that information?

A Yes, ultimately, yes. We have got that, yeah, the final-- yes.

Q Now, this is an opportunity to-- I think we'll do it when we get to the calculations. What I want to do is go to the response document which is bundle 44, volume 5. I'm not going to take you over areas of Dr Drumright's evidence in detail. I'm just going to go to specific points towards the end.

A Sure.

Q If we go to page 50 of this, I think, with luck. Yes. When did you first see this response document and Figure 2f.3?

A So I would say-- I wouldn't be wrong, I hope, if I say in the last month and a half.

Q Right. If, just for the context, we flip between this and 44, volume 1, page 118, so the page we were previously on. Yes. Dr Drumright tells us that, effectively, it's the same data, though there are some adjustments.

Do you have any opinion on

whether the way it's presented in volume 5, if we can go back to that, is of more assistance to the Inquiry?

A Well, I have benefited from the discussion that I had with Dr. Drumright since and I think the conversation-- and of course the fact that Figure 2f.3 gives a lot more detail alongside other paragraphs, 2f.12, so yes. In answer to your question, yes.

I mean-- and if I may, it does so because I agree with Dr Drumright that the generalised additive models we know as GAM models, which have been used by Dr Drumright in the subsequent analysis, do well to help understand-- now I'll use her word, so "wiggliness" in the data.

Q When you say do well what do you mean?

A In that the model that takes account of all the data points or it takes into account the data points to the extent it can. It tries to make sense of it and it acknowledges, because there are more outputs here than in Figure 19 or 22. There we just see one line which is essentially, well, hardwired to the trend.

Q So----

A Whereas here you-- Yes, so it gives you a more sort of nuanced approach, acknowledging the changes in rate-- so at different points in time.

Q I've obviously got Dr

Drumright's evidence on this and she produced it.

A Sure.

Q What do you think is important that we should draw out from this Figure 2f.3?

A What we can draw from that is-- and I'm just looking at the information that has been provided by Dr Drumright, right below, but what we see is that the linear smooth curve in red follows the rise in data points and therefore conveys that from the middle of 2016 to about the end of 2019, or maybe the middle of 2019, that there is an increase in the rate of infection, one that is not captured by the linear blue line.

Q One of the issues I asked Dr Drumright about, she had observed-- I think it's in the text below, if we zoom in on 2014-2016. So if we zoom in, top half of the page, please. There we are.

She discusses in the text the dip in the red line starting at early '14, which, of course, we know to be more than a year before the hospital moves, and I asked her whether she could assist us as to whether that dip is a real artefact of the rates going on at the time or an artefact of the GAM model attempting to get down to the point when there are an awful lot of zeros in late '15, early '16.

I don't think she was able to reach a conclusion. I showed her various HPS

reports that have Yorkhill data. Do you have a view about where we fall on that side of the line or is it, as she says, difficult to tell what's actually going on?

A Well, when you use a model such as GAM, what it is essentially doing is that at every point-- that it wants to lay down the red line, it is accounting for what came before, and so one of the artefacts of GAM models and what makes them useful, and here I'm looking at the linear smooth curve, is that it is quite sensitive to the data points.

So where the linear line sort of takes the line of best fit, so almost ignores the fact that there are the data points in black, because if there are a lot of zeros and the linear line-- the blue line sort of ignores it, going that there seem to be deviations from what I recognise to be where this trend seems to be going.

THE CHAIR: Just from my notes, Mr Mookerjee, I think you said the blue line – in other words, the linear trend – ignores the zero points. Is that what you----

A Well, I-- Yes, I mean, it's a loose term, but yes, I said "ignores", but it's a loose word, but I'm trying to sort of make sense of the-- The blue line, linear lines, are sort of, well, quote unquote-- they're taking the line of best fit. So, essentially, what they're trying to do is--

are there the same number of data points above the line as there are below? And if so, the line needs to be there.

Whereas the smooth function within GAMs, which gives rise to the red line, is a lot more sensitive, so-- and therefore, if you see this clustering of zeros just prior to 2016, well, going to 2016, my take would be that the dip in the red is an artifact of it acknowledging the presence of those zero infections.

MR MACKINTOSH: Right. What I'm going to do is I'm going to take you to your joint work with Dr Drumright. Now, you didn't produce a joint report eventually.

A No, we did not.

Q But you did meet over Teams.

A We did meet.

Q I did say in the paper that was in the bundles that I wasn't going to ask you to discuss the blow by blow or any conversation you had. I just wanted to see whether you could agree.

A Yes.

Q Whilst you didn't agree on a report, you've done some bits of work. I want to look at yours, and then I'll look at hers.

A Sure.

Q So if we go to 44, volume 7 and we go past the opening note to page 37, so you produced this document I think slightly before Dr Drumright produced hers, or was it round about the same time?

A Round about the same time. I think we were both on leave, and then we both came back at different times, and we did one within a few days of each other or a week of each other.

Q I'm going to pass over the Aspergillus questions because I'm going to come to that later.

A Sure.

Q So I wonder if we can go to page 43. I think at 2.7 you start asking questions about BSI bacteraemia rates. Now, did you effectively choose to use her rates rather than yours?

A So in----

Q Or did I tell you to? I can't now remember.

A No, no, I-- to avoid or-- with the time restrictions and the fact that we had a couple of days to be able to do this between leave, so acknowledging the fact that we were working off infection spreadsheets which were different, that--

So I think we both acknowledged that, because of the differences in the source data or the manner in which the source data was presented in terms of the infections, and of course the activity data, that we would expect the rates for the periods that we note here in 2.8, 2.9 and 2.10 will be different. So in order to, well, help the Inquiry, I agreed to take on

the rates of infection calculated by Dr Drumright.

Q Right. That means that if we were to look back either at 2F.3 or Figure 22, it's those points in the charts that you're using.

A Yes.

Q Now, Dr Drumright did hers in a particular way: she used GAM models.

A Yes.

Q She's talked about that, and if we go over the page onto page 44, do we find that you've calculated a series of rates each year or parts of year?

A Well, this was just to understand this. Well, Table 3, I used my initial-- well, the quantitative infection link report on Schiehallion infections, using Dr Drumright's, well, bed days, and I did that as an exercise to understand how far away the bed days that she was using were from the ones that I had access to.

Q So, effectively, this is you just connecting your work to her work?

A Exactly, and it is something that you would do, you know, in the field of epidemiology. You would try to understand how much of a pull is the different activity in terms of the rate.

Q Do you want to draw any conclusion for this, or is it just part of a working?

A No, it only serves to be part of the working because I did not go on to use this for my final analysis where I used the rates which were calculated by Dr Drumright.

Q So over the next page on page 45, you set out your objective to calculate infection rates for three periods. This is January 2008 to May 2015, June '15, September '18, and October '18 to February '22. Do you have any views on Dr. Drumright's evidence about using averages to construct a denominator for the 2005, 2007 period of her Yorkhill data sequence?

A I think it's really, well, problematic, extrapolating bed days backwards. I would draw the line at filling in some seemingly lower than normal or higher than normal bed days with a average, but I would not----

Q So you do the thing they did with the adult haematology rates in the HAD report where they were concerned about some low rates so they put an average in instead. You would do that?

A I would do that, but I would follow that up with making sure that I adjusted the unusually high----

Q Right.

A -- the high ones too because you can't apply it one way and not the other, so I think that's the sad part about the epidemiology. You have to do it both ways. But I would not, based on an understanding-- and that too

retrospectively-- I mean, at the point when the 2008, ’09, ’10 rates were being looked at, we are already in ’25, so our knowledge of what was going on there is scant. And then to extrapolate retrospectively, there’s a huge assumption that nothing at all is different about the hospital in all ’05, ’06, ’07, as opposed to ’07, ’08 and ’09----

Q Right.

A -- which I don’t think we have enough information to say that, so I wouldn’t.

Q Right, let’s go back to page 45 of volume 7. You described your methods. I’d like just to talk through these just to make sure we understand them. What’s the first bullet point saying? What’s reshaping “the data into a long format with a continuous Decimal-Year”? What does that mean?

A So what that essentially means is that, rather than there being a column that is for every month of the year and a role for every year, and therefore every cell is then a year and a month, that what the package allows you to do on R is to essentially go, “Make everything that is Jan of 2015 as 2015.1”----

Q Right.

A -- so that you can lay it out in a long format, and it essentially says what it says on the tin. It’s a long format----

Q It’s just a way of laying out. It doesn’t change the numbers?

A No, no. Yeah, no change of numbers. It just makes it easier to analyse.

Q You then divided the period with three-- with clear time boundaries.

A Yes.

Q Now, Dr Drumright had some concerns about the length of some of the periods that are calculated. Did you have any concerns or have any positive or negatives on that?

A Well, because I work within the clinical setting in Infection Control, I’m very used to adding in, well, boundaries that are based on the clinical need to do so, and we understand these three periods to be these three periods because we understand that the clinical context was not the same within these three periods, and therefore these three periods allow themselves from the clinical perspective to be looked at in three separate sets.

Q Because if I understand Dr Drumwright correctly, her position was something in the broad area of, by doing it this way, you’re imposing your own biases. I don’t think she used that word, but you’re imposing your own perspective onto the data when you should allow the data to tell you where the change points are. Would you agree with that, or would

you disagree?

A I would furnish that with the caveat that, while that might be a good thing to do, that we're working here with-- that when the clinical information allows us to make the distinctions that we can give weight to doing the analysis in that way, in that there's a time and place to be completely removed from the analysis, in that, "Let the package do its bit," and I think we have seen the outputs along those lines.

But there's also a place and time to go, "We understand that these time points are clinically significant. And, therefore, we're almost doing a before/after but comparing these three periods together.

Now, you can do that by doing as we start to do here, by *a priori* looking at the datasets in three separate chunks, or of course, as we can come to later, we can allow the R package to help us understand if, from the perspective of the analysis, are there the changes in rate.

But I think what I'm trying to get at, is that epidemiology within this context exists within the realm of clinical understanding. It is a tool to understand the clinical context better. This is, in my mind, not just the data. These are patients, these are admissions, these are movements, these are infections.

So even though I am an

epidemiologist, I'm not-- I find it uncomfortable sometimes to just call these data points or-- because, actually, they are so much more than that. And so if we need to understand the clinical context of '08 '15 and we want to understand the question about how that chunk in time differs from '15 '18 and then from '18 to '22, we should ask those questions.

Q I think it was Dr Agrawal who, I think in answer to a question of his Lordship, described his colleagues' approach to the material that might have been available but they weren't able to see as being agnostic and focusing on-- and Dr Drumright talked about focusing on - the data.

Am I right in thinking that that isn't a view you would-- Does that have its value, the way they've done it, if I understand it correctly – we've only heard from two out of three – which is to look at the data and see what it finds? Are you taking a different position and saying, "Accept the fact there's a clinical picture and look at that"?

A I mean, surely we can't not acknowledge the fact that we are here speaking about, you know, a hospital or hospitals. We are talking about, well, patients, infections, admissions, morbidity, mortality, so I don't think we can divorce-- or we can't extract out the

clinical sort of-- the need to not have a clinical hat on when we look at these things.

Q All right.

A But saying that, there is a time and place, and I agree that you can run the data in a vacuum, slightly, well, crude way of saying, without acknowledging the context from which have extracted that data. And that has its place, but surely that's all it is.

It'll give you a face value, or there'll be a face value to that analysis, and then the question will arise, well, okay, we have these p-values and we have-- etc., we have the confidence intervals, but does it make sense clinically? What is it telling us? So you can never-- so you always-- you will get back to that question.

Q I wonder if you can go back to page 45 on the screen. This is what you actually did for each period. We'll look at the result in a moment, but you say you "calculated the mean monthly rate" and "fitted the simple linear regression (rate ~ decimal year) to estimate trend slope and p value." Am I thinking that this was a linear regression, not some negative binomial that you were doing?

A Yes, absolutely. This was a regression of the linear variety in line with what I had been doing up till that point, well, fitted to the data points for each of these three sections.

Q Did you calculate these linear regressions separately from the other periods, the three separate calculation (inaudible 15:24:17)?

A Yes, for this piece of work there were three separate calculations, yes.

Q Then what's this second last bullet point? You compared your mean rates for each period using Welch's t-tests.

A So essentially, I guess it's a type of p-value, so it gives you again the significance value.

Q Between two neighbouring periods in time.

A Within the neighbouring periods of time, so, you can compare the mean rates of these three periods to each other and you can infer from it how these periods might be different to each other.

Q Right, and we'll come to the last bullet points in a moment, but let's go over the page and you calculate the rates. Now, this is the Welch t-test calculation rates we see here, or have I misunderstood?

A No, no, the Welch t-test will give you the differences in the mean rates.

Q Right.

A But these are just the mean rates.

Q Oh, these are just the mean rates? Okay, so----

A These are just your----

Q Is this effectively what we see on page 47 graphically, in Figure 2?

A Yes, yes.

Q Right. Let's look at Figure 2 on page 47, so are these the three separate linear regressions you've just described carrying out?

A Yes.

Q Now, are any of those trends, trendlines-- do they have p-values that make them significant?

A The slope, yes. So I think----

Q If you want to go back to the previous page, just say.

A Sure. So the trend for the period June 2015 to September '18 was significant because the p-value for the slope, which essentially tells you how steep or how not so steep the slope is, it is less than 0.05. So it would point to the fact that there was a significant slope upwards here because-- from 2015 to 2018, and then this test----

THE CHAIR: Just almost, really, for my education, the trend figures, what unit is being represented by, for example, minus 0.27?

A So that's giving you the slope.

THE CHAIR: I think I understood that. I just----

A Yes.

THE CHAIR: -- wondered if there is a unit of slope, or----

A I don't----

THE CHAIR: It may be significant to understand---

A Yeah, at this point, but it's a-- but when you sort of----

MR MACKINTOSH: Is it the heading, "Trend per year"?

A Sure, but it is also a value that you can fit into an equation. For example, if you do an equation like Y=MX+C, one of those values is a slope value, but we don't have to get into that. But, essentially, it's-- the slope, well, gives you the relationship between the outcome variable, which is the infection rate here, and the independent variable, which is the year.

So what it tells you is-- and here is where the limitations of the linear model are, but what it's saying is – that for every year between 2015 and 2018, the rate goes up by a value of 3.15 because it assumes that the rate increases by that magnitude for every single year.

Q Linearly, across the period?

A So, exactly, and those we can come to.

Q Just to--

A One of the limitations of the linear model.

Q So the scale is, in fact, 3.5 infections per thousand occupied bed

days for the up, and the minus 0.27?

A I wouldn't put it into those terms.

Q Okay. Right, I won't----

A No. No. I mean, what the line essentially is trying to do is it's trying to give you a value for what links the current infection rate for a year and what it will look like next year, so assuming that, for all subsequent years from that point, rate of infections will rise by a value of-- here of 3.15.

Q Now, have I understood correctly that you're saying that this trend line, page 47, for '15 to '18 is a particular steepness, and it's a statistically significant fit at that steepness?

A Yes.

Q And the green line, the second period, has a particular steepness down but not as steep, and is a little bit less but still significant----

A Still significant.

Q Now, how do you respond to Dr Drumright's observation that use of linear regressions in this context is a little problematic because linear regressions find it hard to take count of all the zeros, and it might have been better, as she did, to use a GAM fit line for these periods?

A So the only thing I would add to that, or how I would amend that statement if I was repeating it myself, would be GAMs are a lot better at

adjusting for the dispersion. That is, you can see-- I mean, if you just, well, focus on the points, you will see a lot of points sort of-- they seem to coalesce, and then you'll see these outsiders.

Q Yes.

A You'll also see that there seems to be, you know, the data points seem to be coalescing at certain points but almost into sort of like a bunch and then they seem to be going upwards.

That GAM models are a lot better-- so, in fact, they are made for being able to better account for the dispersion, in that you understand that, to put it in simple terms, the way in which linear models seem to be going about it, will tend to ignore----

Q So if we were to look at that figure, I recognise that-- and I'm not intending you to draw any conclusions from what I'm about to ask a question, but it might help me understand. Is there a distribution of data points in one of those three periods which is more suitable for GAM models than the others because they seem to have a slightly different distribution?

Can you help us understand? Because you've talked about clumping and you've talked about points that are away from the means and, of course, it would help us, I think, to understand if, for example, you look at the blue section,

that's the Yorkhill section, and then you revisit your explanation of when GAM models would help. Is that an area where GAM models would help, or am I asking the wrong question?

A No. No, I think you're asking the right question. I would just sort of tweak my answer by saying that I think once you-- that there's a process in understanding these sorts of rates whereby you go about to fit the trend line.

So, essentially, you can put a point across each of the lines to look at, you know, how the points go up and down every single month, and you'll find a very up and down-- you know, a trend line which might not suit you because it will be-- it'll be too wiggly, and then through a succession of analysis, you then come down to understanding what model might.

So I think both myself and Dr Drumright at the same point, or-- we agree now, or I agree with her, that GAM models would be suited to look at these infections.

Q In the PSI data set?

A Yeah----

Q We'll come----

A -- because what you're dealing with here is that there is not an overabundance of zero incidence months. I mean, we have zero incidence months, but they don't make up the majority of the infections.

Q That's when you would worry about using them?

A So, for example, with the aspergillus. But what we are trying to take account of is the dispersion in that we just have these points, and we need something to be a lot more sensitive to the peaks and troughs, which I agree with Dr Drumright that GAMs are much better suited for.

Q Well, let's go and look at Dr Drumright's attempts. So we're at page 57 of the same bundle. Now, I'm going to walk through these and then ask you a question. So this is Figure 2.4. This is the GAM model for the Yorkhill period that we asked to calculate.

Dr Drumright decided that linear fit line, neither of them is significant, in the sense that the slight decrease in trend doesn't quite reach statistical significance. So if we go over the page, what did you take that she was saying with this chart, which is 2.5 on page 58?

A Yes, so Figure 5, isn't this not? I'm just trying to wonder whether this is Figure 5.

Q Well, the index-- the labels actually are on the next page, the top of the page. So you're back to 58.

A Well, at least-- well, I was trying to marry up the 0.2.4.2 with this figure, but I don't think-- they might not relate to one another. But saying that,

just sort of looking at that figure, I think the take home, the message from this figure is to take into account the linear smooth red line.

Q Right.

A Because it does a lot better job at acknowledging that late 2016 slight hike in the predicted rate, and what it essentially says is that there is a increasing or that there is a-- I think this red line in her figure was statistically significant.

Q So if we go on to page 60, where she combined the whole RHC period into a single chart, 2.7----

A Yeah.

Q -- and she suggested that the red line linear plus smooth is statistically significant.

A Yes, and this comes back to, and I concur with her point, that one of the requirements of GAM models is that you have to throw, well, as much information you have at them. So they best represent reality, so to speak, when they have more data.

Q This was her point about power and the passage of time -

A Yes, exactly. Because to be able to say something sensible about what the trend infections are and to account for the peaks and the troughs, that doing it over a longer period of time helps, and so, here, of course, from-- and I think she does one for the entire period as well, which is----

Q Which she does. It's the next page, which is the one we've looked at. Page 61. Figure-- it was 2.8. It was 2F3. I just wondered, looking at this chart, to what extent do you disagree with Dr Drumright's interpretation of her own data in this paper that she produced after speaking to you?

A I think the take home message from this graph, so 2.8 originally is Figure 2.F.3, our takes on it are the same, in that the linear smooth line indicates, well, quite clearly that there was an increased risk of infection or there was an increased rate of infection arising, so, looking at the slope, from early 2016, until it sort of meets the blue line. Well, I would think sometime before because of the overlap of confidence in intervals.

So, say, the middle of 2019, that there is a definite peak of infections over that period as compared to the rest of the timeline.

Q Before I ask you about how you might analyse this, to what extent do you see any-- Well, you've actually already answered that, so I won't take up time on that. Dr Drumright discussed the concept of what she called a counterfactual.

A Yes.

Q So I had put to her, or rather

she had been instructed to consider the possible-- what might be seen in the data from a widespread contaminated water system or an inadequate ventilation system. So I asked if she'd mind using the word "scenarios" to describe them, and she did until we put them to one side. Then we talked about counterfactuals.

I think, from memory, she identified issues about line safety, which to some degree related to rooms, single rooms, and nursing resources. Then to the issue around antibiotic prescribing, an issue around laboratory contamination, and an issue around management communication at many different levels.

Now, I think, to be fair to her, she accepts she didn't have a lot of evidence on some of these things. I appreciate that you're not a microbiologist. You will remember, if you nod, the poor person doing the transcript can't write it down.

A Ah. So, yes, I agree.

Q So I'm not going to ask you to give an opinion on the extent to which any of these particular scenarios might be microbiologically sensible or what their mechanisms might be. But I wonder if you can help us understand what you understand a counterfactual is and how to use it as part of the thought processes of thinking about how a peak like this can be explained. So what is a counterfactual, in your mind?

A So a counterfactual, in my mind, in epidemiological terms, would be akin to the word confounder. So, essentially-- other factors that might be working or that might be linked to the outcome variable, which is the infection rate and confounders are something that we have to deal with in epidemiology all the time.

So I'll answer the question in this manner, or you may not have asked this question exactly, but, I'll finish my thought process, so, the first point, we understand that confounders exist within the healthcare environment, right? Because they are fluid environments in that we have a lot of patients, they are in the environment.

We have a lot of other factors. We have antimicrobial prescribing. We have the availability of side rooms. We have staff workload. But these are not unique to a hospital.

I mean, if you go into every single hospital on this land, you will hear that we have issues with staff, you know, the samples of microbiology don't seem to come back as soon as they should, that our prescribing could be better or more patient-centric in terms of the epidemiology or taking more account of the epidemiology of the resistance bugs, that we might have other issues within the hospital.

For example, I know that line-associated BSIs have been noted as one of the confounders. So, essentially, what we are saying is that these are not unique to a hospital, and it's not unique to the RHC. It's not unique to QEUH.

One of the ways in which we adjust for it in terms of when we are comparing rates of infection is by taking into account as large a sample for as long a period we can, because that allows you to bring together an estimate that accounts for the action of these confounders, and essentially weighs it or accounts for it, so across the dataset.

So you, essentially, by accounting for it with large numbers over a large period of time-- so on the proviso that they're all acting the same way in-- so all of these are the settings we are looking at, that we have accounted for them. The few things that I haven't mentioned is-- one of the things I mentioned that was CLABSI the central line associated BSI.

Now, in my report where I provide feedback on the HAD report, I said something which I think should be noted, is the fact that you have bacterial infections that are associated with lines, which primarily are linked to bacteria that are gram-positive, that can exist in the same space as your issue around the environmental infections.

What we need to acknowledge is

that these two things are not mutually exclusive; they can exist in the same space. But what we also have to understand is that gram-positive bacteria are different to those that we are concerned with here in terms of the environmental bacteria.

Q What I wanted just to press you on, I think, a few points, drawing from that discussion, is that I think Dr Agrawal expressed it as clearly as anyone, that the HAD authors were attempting to see if there was a difference in environmental rate infections between one hospital and its predecessors, and other hospitals in Glasgow.

I put to him – and I wonder what you think about this – that the value of using the period before any intervention to address either water issues or ventilation issues or even CLABSI issues, and comparing it with the period after that intervention to see if the intervention made a difference, does that have any value to us to understanding charts like this one on page 61?

A Yes, because you have to go with the hypothesis that we understand that the environmental BSIs which are gram-negative and fungal infections will react to interventions such as what was done on the water system, and so with that hypothesis we go.

We know that we understand from evidence that remedial actions were taken with regards to the water. We know that these, well, environmental bugs reside in the water, that-- So if we do something about the water, what happens to the infection incidents in these environmental bugs? So I think it's very well plausible and a widely well understood epidemiological tool where you do this before/after study almost.

Q Again, I'm going to ask you some questions which I think some of them you might say, "I can't do this, I'm not a microbiologist," so, I mean, I'm conscious of that and----

A Sure.

Q -- ask you to say that if appropriate. If we look at the signal seen in the HAD data of a peak around early 2018, or we look at the signal seen in your work between '17/'18, or we look at Dr Kennedy where it peaks in '17/'18, it doesn't matter which one we're doing, to what extent-- I'm going to use the concept of "some", and by "some" I mean more than a trivial amount but not necessarily all. I'm being quite vague about that deliberately.

From your understanding of the issues so far we've looked at, conscious that you're not a microbiologist, to what extent could some of these environmental paediatric BSI in this peak between '16 and '19 have a cause that's not related to the water system in the hospital?

A I think I would concur that it is likely that some of the environmental infections are not associated with the water.

Q The reason I ask that question is because-- and this is where you might say, "No, I can't do this anymore" - given what we know about how these infections can come from patients, from colonised patients, from the gut and various places like that, to what extent is it a reasonable thought that the rate of those infections would be consistent throughout time? Because same patients, same unit, same practices-- You wouldn't see a peak in those patients? In fact, infections that come from the patients, would they remain consistent? It'd be the ones from the environment or from line safety or from bad management or from lack of-- A transitory issue that would cause the rest. Is that something that we should look into?

A So I hope you wouldn't mind me making this point in relation to the last point where I should have noted that as we, you know-- The assumption that some of the environmental infections might have a source other than water, we would understand this to be the case for all the comparator units as well.

They have also-- well, over this

period of time, as is evident, have had environmental infections. Some of them, I'm sure, would not have a source that is water, and yet the rates or the total rate of infection, environmental bacteria and fungi led, at the Schiehallion is higher for 2017/18.

And I say those two years in particular because there's-- you know, there's a huge amount of overlap and evidence between the analysis undertaken by the HPS, by Dr Kennedy, in the latest analysis by Dr Drumright and in the myriad of things that I've done, because if you assume that something is happening at one hospital, you have to assume the same is happening elsewhere.

And what I'll say is-- on the second point, that the existence of polymicrobial or the extent of polymicrobial sample positives is a cause for concern because what you would expect if for example something was being driven due to or by poor line care, or if you take another confounder, the transmission from one patient to another, you would normally see a bug, one organism.

Q Are you confident this is within your expertise?

A I mean, we can caveat that by saying that it is the words of an epidemiologist who has looked at vast amounts of data and is making this call based on the perspective of his epidemiological expertise, not as a microbiologist.

Q Right.

A And, coming back full circle to Bradford Hill, you know, we have not seen another, well, plausible driver for why the rates at the Schiehallion are so much higher than those of comparators.

Q What I wanted to do now, Mr Mookerjee, was to look briefly at Aspergillus.

A Yeah.

Q Now, it may be we can take this relatively quickly because you weren't originally instructed by the Inquiry to review Aspergillus infections, were you?

A No, I wasn't.

Q Was there a reason for that, as far as you understood it?

A Not particularly. I wasn't privy to the reasons.

Q But, ultimately, you were asked by me to calculate using-- Well, firstly, Dr Agrawal constructed a dataset.

A Yes.

Q I take it you're not in position to criticise his methodology for constructing that dataset?

A No.

Q You and Dr Mumford then used it to construct an incidence rate.

A Yes.

Q And you did that on an annual basis?

A Yes.

Q But you and Dr Mumford added six cases to the case list for the final year, if I understand?

A Yes, for ’22, yes.

Q Do you understand why we might have gone back to Dr Agrawal’s original set to remain consistent with his concept?

A So, again, I think in lieu of the time constraints and to sort of lessen the back and forths and any agreement that can be had about the methodology, more or less the rates of Aspergillus year on year were fairly similar between what was calculated by Dr Agrawal and by myself and Dr Mumford. So we thought that we could just go with the rates calculated by Dr Agrawal.

Q Yes, and I just wonder whether at the end of this journey of then Dr Agrawal, Dr Drumright and Professor Hawkey calculating GAM fit lines for adult and paediatric Aspergillus, and then you and Dr Drumright speaking about the matters and she constructing GAM fit lines and you constructing linear fit lines-- If we go to the HAD report, page 128, that’s volume 1 of bundle 44, and this is the new numbers that Dr Agrawal created----

A Yes.

Q -- from which the incidence rates were created. Where do we get to about those incidence rates? In your opinion, is there any statistical significance in the suggestion that the rate of Aspergillus infections before the move to RHC is different from the rate afterwards? Do you have a view on that now?

A Yes, so-- And my view is based on what confronts us is that the rates of Aspergillus are based on very few numbers in terms of infections. By “very few”, it’s a word that I use within the context of what we see with the other BSIs. I mean----

Q Yes, so in this particular case, for the children, it’s 0 to 5 at Yorkhill and 1-7 each year in RHC.

A Yes, so the only thing that you can say is that there is a wider spread in terms of the infection incidence at RHC as compared to Yorkhill. As an epidemiologist who works within the infection control setting and has done, looking at this chart with my clinical epidemiology hat on, I would-- at my workplace, this would flag a need to look into the numbers and to look at what is driving the increased incidence because the trend matters.

And if you have a line, well, connecting each of the-- at the top of these bars from 2008 to 2022, you would

see that that curve sort of will make its way up from somewhere from 2 and you start seeing a lot of 5 and 7s and 6s.

Q Yes.

A Because within the field of clinical epidemiology, we don't wait for something to be to be significant in terms of the p-value to be able to do something about it. You have to be quite reactive to the problems that are happening, and even in retrospect, this would definitely flag as cause for clinical concern. Not everything has to be backed up by p-values.

Now, saying that, since we are on the subject of what models that we have used so far can tell us about how different the rates of Aspergillus are between Yorkhill and RHC, my view is that the numbers are too few or the sample size is too small. There are a lot of zeros, so there are a lot of months within this period where the count of Aspergillus is zero.

For any models to really make anything of this-- so infection rate over time, in that we can't say one way or another whether there is a actual change----

Q In the paediatric patients?

A In the paediatric patients. I haven't looked at the adults in the same amount of detail, so I'm just focusing here on the paediatric. Saying that, you can see from the case incidence that there is an increase in the incidence numbers for Aspergillus that I don't think can be accounted for by an increase in activity, and therefore that says something to me broadly about--

Or I would ask myself the question, are more patients returning positive-- with a positive sample for Aspergillus? And how I would answer that is, more or less, "Yes, between or after-- so 2018."

Q Okay. I think, Mr Mookerjee, that I must-- I think I have asked you all the questions I had planned to ask, but what I would like to do, my Lord, if possible, is to take 10 minutes to confer with my colleagues in the room and those watching remotely for them to see if there are any further questions for Mr Mookerjee.

THE CHAIR: As you may recall from your last attendance, Mr Mookerjee, the process we are following is to allow Counsel to check with colleagues to see if there are any more questions that should be asked, so it should take no more than 10 minutes. If I can ask you to return to the witness room.

A Sure.

(<u>Short break</u>)

THE CHAIR: Mr Mackintosh.

MR MACKINTOSH: Yes, my Lord, I

have one question.

THE CHAIR: One question, Mr Mookerjee.

MR MACKINTOSH: Yes. Mr Mookerjee, I wonder if we can go back to the document you produced over lunchtime.

A Yes.

Q So it's been suggested to me that in the second row for 2016, if one divides 25 by 2,226 and you divide that number by what happens if you divide 72 by 14,714, the comparator rate, you don't get 2.46, you get 2.25.

A So just to point out that when you do this calculation, it's not like as if I'm dividing 72 by 14,714. It takes into account-- because you want to add the raw data to the analysis, so it's taking into account all the proportions, so the four proportions that make up the overall rate of 4.5. Do you see what I mean?

Q So you're saying it's not as simple as that?

A It's not as simple as-- I mean, these are given for-- was one of the reasons why I didn't include it in these terms is because if you look at GGC, there's one set. There's one proportion, so to speak. It is 25 divided by 2,266 because there's no other data. There's one unit.

Q Yes.

A But this 72, for example, is made up of four separate numerators, and 14,714 is made up of four separate activity figures, and those are all taken separately. So if you can just imagine----

Q So they have a weighting effect, in some degree?

A Well, they will not exactly be equal to 4.5, and so the rate ratio is not just a difference between 11.03 and 4.5.

Q Okay. Then in respect to 2022, the rate ratio shouldn't be 0.65; it should be 0.64.

A Yeah, again, it's a facet of-- it's not 89 divided by 18,655. It's each of those four numerators, and each of those----

Q I think what we're going to do, Mr Mookerjee, is we've asked Core Participants who wish to express a view about this particular aspect of your evidence----

A Yes.

Q -- and in particular the statistical significance of the rates in '16 and '17 and '18 to do so by email to the Solicitor to the Inquiry in the next couple of weeks. We'll then work out what to do with that, and we will pick that up before a particular person gives evidence in, I think, Week 2 of Part 3. My Lord, with that, I have no further questions for this witness.

THE CHAIR: Right. Mr Mookerjee, you are now free to go----

A Thank you.

THE CHAIR: -- but you go with my thanks for your attendance today, your previous attendance, and the considerable amount of work that that has involved, so thank you. You're free to go. Thank you, Mr Mookerjee.

A Thank you.

(<u>The witness withdrew</u>)

THE CHAIR: Well, we shall, all being well, see each other tomorrow, ten o'clock, with----

MR MACKINTOSH: For Professor Hawkey.

THE CHAIR: -- Professor Hawkey.

MR MACKINTOSH: Yes. Thank you, my Lord.

16:30

(<u>Session ends</u>)