Page 1

Friday, 27 September 2019

(10.17 am)

CHAIRMAN: Yes. Thank you.

Firstly, could I apologise for keeping everybody waiting. I thought that the matter might take just a minute or two before 10 am, but it took somewhat longer than that.

The fact of the matter is that there's a particular point which is not one of evidence, it's one of the way forward, and it is actually a point of some fundamental importance, in my view. I may be wrong, but I feel that the matter by way of the Commission is worthy, at this stage, of a brief discussion with all counsel, in chambers, just to try and understand the way forward; all right? I'll explain the position and seek your assistance.

We were thinking of perhaps proceeding more formally by way of letters, questionnaires and the like and then having a meeting maybe next week, but to be honest with you I feel that the sooner we get to grips with this, perhaps, the better. All right? It may well dictate how we proceed in the future. It may well save us time, and therefore cost.

That all sounds a bit intriguing, no doubt, but what I would like to do is just adjourn, shall we say --

Page 2

well, we'll just adjourn and then I will speak to everybody in chambers, explain the concerns that we have, and then we can get an explanation in chambers, off the record -- it's got nothing to do with any commitment of evidence; it's all to do with how best to proceed, that's all -- and then we can return and we can hear from the professor.

Professor, I'm very sorry. I appreciate I'm keeping you. You got here on time this morning. My apologies for that. But hopefully we will be able to continue with your evidence and complete it today. All right? Thank you very much indeed.

So where would be the best place to meet? Because if it's in my chambers, it's going to be --

SECRETARY: The transmission room.

CHAIRMAN: All right. So we will meet next door in two or three minutes. Thank you.

(10.21 am)

(A short adjournment)

(11.18 am)

CHAIRMAN: Just before we start, I think, for the benefit of the public, I should say just a couple of words, namely that myself and Prof Hansford have met with counsel and instructing solicitors in chambers, for the purposes of considering the best way forward for this Inquiry,

Page 3

having regard to some fundamental matters which may well dictate the way forward in a way that is aimed at saving time and still meeting the terms of reference of the Commission.

That sounds very general, and apologies to the public, but what I can say is that what was discussed and what will still be discussed, in order to better fashion the way forward, will be formalised and made public in the next few days. So nobody is keeping anything from the press or from the public, but, as always in cases of this kind, initial ideas, initial concerns, have to be better formulated, they have to be discussed, everybody has to have some commonality of purpose, and we have to make sure that the terms of reference are being honoured, and once that is done then everything will be made known and both the press and the public will have a far better idea of where we are going.

So this is one of these situations where we just need to get our house in order and then we will open the door of that house to the press.

Thank you.

MR KHAW: May it please Mr Chairman, may I now call the government's statistical expert, Prof Yin Guosheng.

Page 4

PROF YIN GUOSHENG (affirmed)

Examination-in-chief by MR PENNICOTT

Q. Thank you, Prof Yin.

We understand that for the purpose of this Commission of Inquiry, you have submitted two reports, one report for the Original Inquiry and one report for the Extended Inquiry.

If we can take a look at the two reports. The one regarding the Original Inquiry, it's item number 12, ER item number 12; can you see that? The hard copy and also the soft copy should appear in front of you.

If I can just take you to your first report. You see your name at the top and also there's a signature on page 2?

A. Yes.

Q. You confirm that that's your signature?

A. (Nodded head).

Q. Then the report actually consists of various pages. It goes all the way --

COMMISSIONER HANSFORD: Sorry, I think Prof Yin needs to be told that we can't take nods because they don't end up on the transcript, so you have to say "yes" or "no", so it ends up on the transcript.

A. Yes. Okay.

MR KHAW: For the purposes of the record.

Page 5

Thank you, Mr Commissioner.
You see that the report consists of various pages and it goes all the way to page 20.
A. Yes.
Q. Then, after page 20, you see that you have also given us your CV?
A. Yes.
Q. First of all, you confirm that the contents of your CV are true and correct?
A. Yes.
Q. If we can then take you to the second report, that is for the Extended Inquiry, ER item number 4. Do you see that?
A. Yes.
Q. Again your name appears at the top of page 1?
A. Yes.
Q. And there's a signature at page 3; do you see that?
A. Yes.
Q. You confirm that that is your signature?
A. Yes.
Q. Now, in relation to these two reports, insofar as they contain factual matters --
A. Yes.
Q. -- that you have outlined, you confirm that those factual matters are true and correct?

Page 6

A. Yes.
Q. Insofar as they contain your opinions, do you confirm that they are your true and honest opinions?
A. Yes.
Q. You have also prepared a response to Dr Wells' report, and in fact that has been uploaded to the bundle and it can be found at ER4.1.
A. Yes.
Q. Do you see that?
A. Yes.
Q. May I also confirm that your signature appears at the first page of that response?
A. Yes.
Q. And it consists of several pages, up to page 16; right?
A. Yes.
Q. So you also confirm that the contents actually contain your honest and true opinions?
A. Yes.
Q. I understand that you have, for the purpose of today, prepared a synopsis.
A. Yes.
Q. I think you will be shown the synopsis on the screen.
A. Okay.
Q. Perhaps I will now leave it to you to present the points contained in your synopsis.

Page 7

Oral synopsis by PROF YIN GUOSHENG
WITNESS: Okay. Thank you very much. Mr Chairman, Mr Commissioner, good morning, everyone, I'm very honoured to come to this place to share with you my statistical analysis about the whole investigation.
My name is Guosheng Yin, I'm a professor and also the head of the department of statistics and actuarial science at the University of Hong Kong.
Next page, please. The question about this coupler connection, whether it's defective or non-defective is simply a "yes or no" question. So it's just like tossing a coin. You observe a head or observe a tail. So this kind of random variable follows what we call binomial distribution, and the equation in red here is binomial distribution probability maths function.
So we are interested in estimating p, so it's defect rate in the whole structure, and the sample size is n, and the y is the number of defective coupler connections in the sample.
Once we've estimated p, let's call it p-hat, based on our sample, and then we can construct a 95 per cent confidence interval, and it's given by the second equation on the screen, right here (indicating).
So we got this 95 per cent confidence interval, it basically shows at the bottom curve, you can see on the

Page 8

left side and on the right side, they are 2.5 per cent. That's basically the two tails of this bell-shaped curve, and in the centre is 95 per cent. So, basically, this is the most commonly used statistical confidence interval, trying to characterise the variability of your estimator, which is called p-hat.
Next, please. But for this purpose of this investigation of this defective coupler rate, we are only concerned with the upper bound of this confidence interval. That's basically I call "pu", represents upper bound. So instead of we use a two-sided confidence interval, we use one-sided, because we are only concerned with upper bound of the defect rate.
Next slide. So the first question we ask how many samples we need in order to have an accurate statistical estimation. So, basically, the sample size estimation problem and --
CHAIRMAN: Can I just interrupt one thing?
A. Yes.
CHAIRMAN: We are talking about defects, and defects, I take it, are described numerically for you by the people who instructed you, in other words, what constitutes a defect.
A. Yes, we are talking about defects.
CHAIRMAN: And what is a defect?

Page 9

A. That's a definition from engineering.
CHAIRMAN: That's right. So the engineers have given you a definition of what they consider to be a defect --
A. Yes.
CHAIRMAN: -- and you work from that?
A. Yes.
CHAIRMAN: Thank you. So you yourself have not determined what is a defect in the first instance?
A. I don't. Thank you.
I'm a statistician, so my job basically is once the data being presented to me, I will carry out a statistical analysis. I don't define what is called defects.
So, at the designing stage, there is no information about how the defects would be, like what is the defect rate in the structure.
So what we did is we applied this binomial probability, the same formula you saw in earlier slide. So basically we try to characterise pu, which is upper bound of the 95 per cent one-sided confidence interval for the defect rate, and n is the sample size. So we need to estimate how large n is. In order to estimate how large n is, we need to consider different scenarios.
So basically -- next slide, please -- so we considered different scenarios for y versus pu. So

Page 10

given the sample size of n equals 84, then if you observe zero failure in the sample, then that gives you the maximum failure rate at 95 per cent confidence level, 3.5 per cent; okay?
If you observe one failure in the sample, then that maximum failure rate would go up to 5.5 per cent. So it continues with the number, number of failures observed in the sample, you can see on the right column the maximum failure rate in the population actually continues to go up.
COMMISSIONER HANSFORD: Sorry, Prof Yin, this is based on the sample size of 84?
A. Yes.
COMMISSIONER HANSFORD: Which you have already determined --
A. Yes.
COMMISSIONER HANSFORD: -- by the previous method, as you explained on the previous slide?
A. Yes.
COMMISSIONER HANSFORD: So, once you've determined that the sample size is 84, you are now telling us this is what finding a number of failures in the sample means in terms of failure rate in the whole population?
A. Yes. A very good question. The reason I say that is because we explored another sample size too, for example 50 or 100 or 200. So the whole sample size estimation

Page 11

process is you need to -- I have to discuss with engineer what is kind of practical number, because if the number is really, really high, it will endanger the whole structure.
COMMISSIONER HANSFORD: Of course.
A. So, in the end, 84 is what we concluded, the sample size 84 is for each slab.
Next slide, please.
COMMISSIONER HANSFORD: Sorry, so 84 was determined as being the optimum number or the minimum number, was it, for the sample size? Because you said you determined -- you looked at 50, you looked at 100 --
A. Yes.
COMMISSIONER HANSFORD: -- and you determined 84 to be optimal?
A. Yes. Let me tell you why we decided 84 eventually, because if you look at the first row, if zero failure, then the maximum failure at 95 per cent confidence level is 3.5 per cent.
COMMISSIONER HANSFORD: Yes.
A. And this is low enough. What I mean by "low enough" -- because in statistics, we often have this 5 per cent significance level, or 5 per cent we consider is kind of threshold. So this is below 5 per cent, so we think if there's zero failure, then 3.5 per cent is a good

Page 12

number.
COMMISSIONER HANSFORD: Whereas, if you had used 50, it would have been higher than 3.5 per cent?
A. Exactly, yes.
COMMISSIONER HANSFORD: Okay. Yes.
A. I think in my report, I had it.
COMMISSIONER HANSFORD: You did.
A. But I forgot exactly the number.
COMMISSIONER HANSFORD: I'm just trying to refresh my memory on the point.
A. Thank you.
Next slide, please. So we implemented a two-phase sampling scheme. Phase 1, we randomly select the D-wall panels and also the location within the selected panels.
So, at the bottom figure, you can see there are many, many panels, and they have various lengths, from 2.8 metres to 7.2 metres, and the width is constant at 1.2 metres. But the opening-up area is about 0.4 metres, so it's basically a small square, so that you can expose three couplers.
We also discussed: can we expose one coupler at each opening area? And basically the conclusion is you have to open a certain size so that people can go in to do the measurement. You cannot just have very tiny crack and then to do all the measurement. So eventually the

Page 13

conclusion is it has to be a certain size to expose three couplers.

Next slide, please. So once the panel is selected, the panel you look is a blue colour in the figure, and the panel is quite big, so we need to determine where exactly we want to open this 0.4 metres square, and that basically we need to generate another random number so that we can select the location. And once we select that location, we will open that small square, try to expose three couplers.

Next slide, please. Phase 2. Once you identify the site, there are multiple layers underneath the concrete. It varies from one to five, but here just a cartoon to show you, illustrate how the layer is going to be selected randomly.

Suppose you open -- you select, randomly select the level as a third layer. That's basically "site K" on the far right-hand. I use the red colour to denote selected layer. Suppose you select the third layer, so people need to go there, open that area up, and it has to expose the top two layers which are in colour green, and those top two layers will be opened and will be measured too. So, in a sense, even though our original sample size was 84, but in the end we could end up with a larger sample size because those extra layers, you

Page 14

have to open it in order to go to the third layer.

Next slide, please.

COMMISSIONER HANSFORD: Sorry, does that introduce any bias?

A. No, it wouldn't introduce any bias, because all these are being randomly selected. It's just sort of an extra or bonus sample size goes into -- you know, sample size estimation is always an estimate.

COMMISSIONER HANSFORD: Yes.

A. Just as in any study, clinical trials or any study, you need to determine how large your study will be, and that -- nobody can give you accurate: you have 100 people, you have to have 100 people --

COMMISSIONER HANSFORD: I understand that, but I just wondered, by having additional samples in certain locations, does that introduce a bias?

A. No, I don't think so.

COMMISSIONER HANSFORD: Okay.

A. Next slide, please.

So whether this will result in a genuinely "random" sample -- I've just gone through this two-phase sampling scheme. So we basically use the three randomly generated numbers to select each set of three couplers, and you imagine this is the three-dimensional space you are trying to draw random samples; this is not a single line. If a single line, you can just use one random

Page 15

number to identify the location. If it's a two-dimensional plane, you need two random numbers, but now you are in the three-dimensional structure, so you really need to think three-dimensional space.

So, basically, we randomly select the panel, we randomly determine the reference point, that's the location where you want to open it up, you randomly select the layer.

So I have seen Dr Wells' report. He mentioned whether it's random or not -- first, in his calculation, he treated panels as like couplers in terms of numbers. I will come back to this point later on. But then the bottom, in red, the bottom point: although each panel is supposed to contribute three couplers, as I mentioned earlier, if you happen to select a deeper layer, then there would be extra couplers come out.

So, basically, even though we intended three couplers per panel, but some panels could have six, some panels could have nine. So that's the point I want to say.

Next, please. So once I got this data, I did this plot. Basically, the y axis is engagement length, the x axis is panel number. You can see "EWL" on the left, "NSL" on the right, and for each panel number, sometimes you could have multiple dots, you could have six dots,

Page 16

you could have three dots or even less than three dots, sometimes two dots.

So this is the data I just want to present to you. What I want to really focus here is, if you look at the bottom at EWL panel, there are eight dots have zero engagement length, and on the right side, NSL, there's no zero engagement length.

Next, please.

COMMISSIONER HANSFORD: Sorry, sticking with that for the moment, did you do any -- did you derive a mean from that? Did you analyse all these data to look at what the mean was --

A. Yes, I can.

COMMISSIONER HANSFORD: You can?

A. I don't exactly remember but it should be around 35 millimetres, because it depends on which panel you talk about. EWL, the mean is lower than the mean of NSL, obviously, because those zero engagement lengths.

COMMISSIONER HANSFORD: Yes.

A. So it's a very simple calculation. You can derive the mean.

COMMISSIONER HANSFORD: Yes. Okay.

A. So this is just gives you a graphical look at the data.

Next slide, please. Let's focus on these unconnected couplers. So, in EWL sample, eight out of

Page 17

90 couplers have zero engagement length. So, using the formula I showed you earlier, we can come up with 95 one-sided confidence interval upper bound for unconnected coupler rate is 15.5 per cent. This is only using EWL slab data; okay? So this is basically unconnected coupler rate can be as high as 15.1 per cent. And in the data, we have seen that some of the unconnected rebars only have one to two or three to four threads, so clearly that's indicating some threaded ends were cut.

So let's come back to the engineering criteria for passing. So this criteria, basically, engineers' definition: (i), you have a maximum of two full threads exposed; (ii) -- this is "and", I put the "and" in blue, so you must satisfy these two conditions simultaneously -- engagement length of the threaded steel bar inside the coupler should be at least 40 millimetres, given there's tolerance of 3 millimetres for PAUT, which is ultrasonic technique, that's also an engineering issue, the equipment, the reading below 37 millimetres would be regarded as defective.

And the third bullet is you don't use PAUT. You basically direct measure. You have to have at least 40 millimetres' engagement.

So that's the criteria given by the engineer

Page 18

definition.

Next slide, please.

COMMISSIONER HANSFORD: Sorry, so on this, you are telling -- your understanding on the engineering criteria is the passing criteria is (i) and (ii)?

A. Yes.

COMMISSIONER HANSFORD: Or (i) and the item underneath the line?

A. No, not "(i) and". The third bullet underneath is a gold standard, because that's actual. They really unscrew the bar and measure.

COMMISSIONER HANSFORD: So, if you do the one under the line, you don't need (i) or -- you don't need (i)?

A. Yes. The bottom one is direct measurement, it's the gold standard.

My understanding is -- I'm a statistician -- they don't want to open every one. They want to use ultrasound. Because if they open everything they basically destroy the coupler connection, so they try to use ultrasound to do the measurement --

COMMISSIONER HANSFORD: Which is point (ii) on here, not point (i).

A. Yes.

COMMISSIONER HANSFORD: Which is why I asked whether an option was (i) and the item below the line.

Page 19

A. Perfect, yes, exactly.

COMMISSIONER HANSFORD: So it is?

A. So, no, it's (i) and (ii), this is passing criteria if you use PAUT.

COMMISSIONER HANSFORD: Otherwise it's just "not less than 40 millimetres".

A. Yes, direct measurement.

Next slide, please. Based on this passing criteria --

COMMISSIONER HANSFORD: I'm sorry to keep interrupting you.

A. That's fine, I like questions.

COMMISSIONER HANSFORD: You have referred to a gold standard several times. Well, you have referred to it today and it was referred to in reference to your report yesterday, as thought that's a defined term. Is a gold standard a defined term?

A. Okay, can we go to the previous slide? The third bullet there, that's engineer definition.

COMMISSIONER HANSFORD: Yes, but you said "gold standard". I'm just wondering why you call it "gold standard".

A. Because this criteria would possibly override the other two. That's my understanding.

COMMISSIONER HANSFORD: So does "gold standard" mean overriding?

A. In this case, you can think that way.

Page 20

COMMISSIONER HANSFORD: I'm just asking you what the term "gold standard" means.

A. "Gold standard" has different definitions. For example, if you have high blood pressure, then there's a threshold, defined by the medical doctors, 130 above, you will have hypertension; below -- so this is the gold standard. I don't want to over-interpret the gold standard here. This is just the rule people use, the common rule people use. That's my interpretation.

But this slide basically has nothing to do with me. This is all --

COMMISSIONER HANSFORD: No, I'm not talking about the criteria. I'm talking about the term "gold standard", what the term "gold standard" means.

A. You mean statistically? Because I'm a statistician so I would interpret it from a statistical perspective, and if you are an engineer then you would have --

COMMISSIONER HANSFORD: It's just -- I don't necessarily need to labour this point, but I've not seen in any documentation a statement, "The gold standard will be this", so I'm just wondering where this term "gold standard" has come from.

A. I can give you a simple example. For statistical hypothesis testing, you have a null hypothesis versus alternative hypothesis. You calculate, test the

Page 21

statistics, you obtain a p value. If the p value less than 0.05, we reject the null hypothesis. 0.5 I consider gold standard. Basically, it's the criteria everybody uses to reject the null hypothesis.

COMMISSIONER HANSFORD: So "gold standard" means the criteria everyone would use?

A. Everyone would accept that, yes. That's a typical thing that's commonly used, most commonly used criteria. That's my interpretation.

COMMISSIONER HANSFORD: That's fine. We'll move on. Thank you.

A. So based on the passing criteria in the previous slides, here's the result. The EWL, we have a total of 102 samples, but only 90 give valid results. So 90 have valid results and, out of 90, 25 are defective. In NSL slab, a total of 99 samples, and out of 99 samples there are six missing data, so you end up with 93 observations, and of this 93, 23 are considered defectives. Based on this data, the bottom two lines in red, for EWL defect rate, this upper bound 95 per cent confidence interval was estimated to be 36.6 per cent. For NSL, defect rate was estimated to be 33.2 per cent.

Next slide, please. So, almost near the end of the opening-up exercise, a new situation arose. So we were told that there are capping beam on the D-wall side and

Page 22

there are couplers -- you can see there are couplers that are being exposed both left and right sides. So this is a situation that is unforeseen. When we design the whole sampling procedure, and then this situation arises, what we are going to continue to do -- re-formulate our statistical analysis.

COMMISSIONER HANSFORD: On your diagram there, you show couplers on the left-hand side. You don't show couplers on the right-hand side.

A. Which diagram?

COMMISSIONER HANSFORD: On the slide here that we are now looking at.

A. The left side or the right side?

COMMISSIONER HANSFORD: Sorry. What I'm looking at is -- or the left side, I've got blue.

A. Yes.

COMMISSIONER HANSFORD: And I can see couplers.

A. Yes.

COMMISSIONER HANSFORD: Where on the right side do I see couplers?

A. No. The yellow part is capping beam.

COMMISSIONER HANSFORD: Correct.

A. And below is the D-wall --

COMMISSIONER HANSFORD: So you are talking about the vertical couplers coming out of the D-wall?

Page 23

A. No, no, no. I'm not talking about the vertical couplers. Sorry.

COMMISSIONER HANSFORD: I'm confused.

A. Can you go to the next slide, please. So here (indicating), can you see this blue side, blue colour?

COMMISSIONER HANSFORD: Yes, of course.

A. That's the red coupler I'm talking about.

COMMISSIONER HANSFORD: No, the couplers are -- sorry, where are the couplers?

A. Right here (indicating), the red, you see the arrow pointing down. Right here, can you see on your screen, this one (indicating)?

COMMISSIONER HANSFORD: I've got those two, yes.

A. That's the coupler I'm talking about.

COMMISSIONER HANSFORD: But you said they are on the left side and the right side?

A. I'm talking about -- okay, then you come to the zoom, the zoomed picture. I'm talking about the coupler connection on the left.

COMMISSIONER HANSFORD: I understand. So you are referring to just one coupler --

A. Yes.

COMMISSIONER HANSFORD: -- but you're talking about threaded bar on the left side of it and threaded bar on the right side of it?

Page 24

A. Exactly, yes.

COMMISSIONER HANSFORD: Now I understand.

A. Sorry, I may have confused you. So that plot is just a zoomed-in plot.

COMMISSIONER HANSFORD: I understand.

A. Next slide, please. So, for coupler, you have both left and right side, you need to consider both sides are properly connected --

COMMISSIONER HANSFORD: Yes.

A. -- in order to consider the whole coupler is sound. So, based on that argument, you can see, for the capping beam section, our sample size is 11, so we have 11 such couplers, and on the slide set, there are seven observations. That means there are four missing data.

Out of these seven observations, there are two defectives, and on the capping beam side, we don't have missing data, so you have 11 coupler connections, and out of 11 you have two defectives.

So, based on this capping beam section, we calculated the upper bound of 95 per cent confidence interval defect rate is 68.3 per cent.

COMMISSIONER HANSFORD: But the sample size is rather small.

A. For the sample size of 11.

Let me explain why you can have such high defect rate. I give you a very intuitive explanation, not

Page 25

rigorous calculation. The rigorous calculation is given in my report. But here I just want to explain intuitively, loosely speaking. We look at the couplers on both sides. There are four possibilities. Left/right, both pass; left pass, right fails; left fail, right pass; or both fail. So four possibilities. A coupler would be considered to be a sound coupler only for pass/pass. And that's why I have this bottom equation. You consider our previous calculation for EWL, the defect rate is 36.6 per cent. That is the defect rate only for one side.

Let's consider this defect rate also be used to the other side. Just presume on both sides you have the same defect rate. Then the bottom equation basically gives you -- it's not rigorous calculation but gives you some sense that if you consider both passes as a pass, then you would have 1 minus, the probability of pass, times the probability of pass, on left and right sides. That would be a fail. So that's a probability of failure.

So, based on that kind of argument, you can see why I could end up with 68.3 per cent.

COMMISSIONER HANSFORD: But I'm seeing 59.8 per cent there.

A. That's very non-rigorous calculation, just for intuitive understanding, but in the detailed calculation it's much

Page 26

more complicated than that.

COMMISSIONER HANSFORD: Okay. I understand.

A. So the number I want to report is 68.3 per cent.

Next, please. As Mr Commissioner you have questioned the sample size is rather small, whether the calculation is reliable -- I totally buy the argument. So we went back to use another method, which is called bootstrap. Bootstrap is a widely used method to calculate the variance. So, if you look at the bottom row on the table, if you apply bootstrap method, which is different from what we originally used method, which actually Dr Wells is concerned with because he was thinking the sample size is small, you are doing normal approximation and you are using Delta method, that's fine. So we come back, we do the analysis using another method, called bootstrap, and you can see the estimate is rather very close.

Next slide, please. So there's another concern about why there's no samples in area A. So here's my explanation. Some of the EWL top panels use through bar -- no couplers, no need to sample them, so they are excluded from the random draw. The EWL bottom panels use couplers, but area A are blocked by mass concrete infill between the two slabs, so the couplers are not accessible and they were excluded from the random draw

Page 27

too, but not all of them. In fact, two panels in area A were included in the random draw, but none of them were chosen by our random sampling procedure. And the list of panels with couplers, after you remove all those panels and the through bar are not accessible, the list of panels were provided to me by engineers.

Next slide, please. So here, just to show you area A, where is the mass concrete -- it's being circled in the left, that's basically mass concrete, and that caused inaccessibility of those couplers.

Next slide, please. So this illustrates the random sampling procedure. Indeed, you see we have the third column, "Works area", you can see "HKC" and "A", that's area A. They do have two panels in area A were put in the draw but it was not selected, by random chance.

Next slide, please. Here is another question, whether we can extrapolate the estimates based on HKC to area A. First, I was given that HKC and area A with capping beam have very similar configurations; same contractor, so similar workmanship. In HKC, we had 11 samples. At the capping beam side, there were two failures. In the EWL slab side, there are seven valid observations, because there are four missing data. Out of those seven valid observations that have PAUT results, two failed. That's why we use this data,

Page 28

combine both left side and right side of one single coupler, to calculate the combined defect rate, so we get this 68.3 per cent.

Okay, now the question is whether this result can be extrapolated to area A. I think this question is more of an engineering question, but from a statistical point of view I don't see why not. I just leave it.

Next, please. So Dr Wells back-calculated, trying to check whether our sample is a genuinely random sample. He used 175 versus 62 panels. He considered that's like a population of panels, versus 83/7 sample, but those numbers are couplers. So, basically, he was trying to calculate the ratio between no capping beam panels and capping beam panels, but he was thinking panels and couplers are the same thing. Okay? Then he recalculated the whole thing in proportion, because easier to visualise proportion. So basically he said in the panel there are 26 per cent capping beams, and in the couplers there are only 8 per cent, and my understanding, based on his report, he had used Z-test for sample proportion with normal approximation. That's my understanding, as a statistician, of what he has done.

So although three couplers were intended to be chosen for each selected panel, but as I said earlier

Page 29

some deeper layers would inevitably expose those upper layers' couplers. So you basically could have more couplers than what you intended to get.

The last bullet is we do not expect the proportions to match between the population and the sample.

Next slide, please. So let's redo the calculation, use what Wells has argued: a total of 237 panels out of 62 panels with capping beam, actually there are only 29 panels in the draw. Out of 175 panels without capping beam, there are only 168 in the draw. So the proportion actually should be 29 divided by 29 plus 168, which is 14.7 per cent. I'm using his argument, without distinguishing panels or couplers, just for the sake of argument.

So we removed a lot of panels because either no coupler, they use a through bar, or they cannot be accessed due to the mass concrete. I think Dr Wells may not be aware of all these details so he couldn't do a similar calculation. But anyway, a total of 102 couplers were sampled for the EWL slab. Among them, 11 with capping beam. So what we should have done or what Dr Wells should have done is use 11 divided by 102, instead of using 7 divided by 9. The reason is we originally sampled 102 couplers due to their 12 missing observations, so 12 couplers were removed because we

Page 30

only have 90 valid observations.

So when you try to challenge whether a sample is genuinely random or not, you cannot do the data processing and then compare the proportions. You have to use the originally sampled coupler, that's 102, and 11 out of 102 has capping beam, so you should use 11 divided by 102, which is 10.8 per cent.

So the red line there I point out, "Should not use the numbers after removing the missing data", because after you remove the missing data, you have basically already done data processing.

Next, please. There's another point --

COMMISSIONER HANSFORD: Sorry, can you just repeat that point you just made? After removing -- or does that come up later? You said after removing the missing data, you have already done data processing. I don't understand.

A. Basically, you think about it this way. You have a population, you want to get a sample, and you want to see the population proportion and the sample proportion are comparable.

COMMISSIONER HANSFORD: Yes.

A. That's his whole argument. But in the sample, you should use the originally randomly drawn sample regardless of data missing or not. We are not talking

Page 31

about the engagement length or PAUT. We just say, "How many samples you have drawn and what is the proportion with capping beam versus without capping beam"? You shouldn't delete those data first, then use that number to compare with population. Is that clear?

COMMISSIONER HANSFORD: Yes. The bit I didn't understand was you said because you have already done data processing --

A. When I talk about data processing, I'm talking about mean removing the missing data. That's all I mean.

COMMISSIONER HANSFORD: That's fine. It sounded more technical than that. Thank you.

A. Okay.

Then Dr Wells' report, paragraph 4.7, he mentioned "a major reason for defects is poor workmanship, then defectives will probably be in clusters, and therefore not independent". Then he argued: because it is not independent, this would lead to higher rates of defectives in the sample than in the population, so any results will necessarily be more conservative than should be the case.

So I will visit the first point first, the paragraph in red first.

Next page, please. So I have done a permutation test to check the independence assumption. I don't want

Page 32

to go into detail the four bullets below the table, but I just point out the table, the meaning of the table.

Basically the three numbers are p values. We are doing hypothesis testing, to test whether the sample has clustered or independent. So, for EWL, p value is greater than 0.05, so that basically indicates there is no clustering, at 5 per cent significance level. NSL is below 5 per cent; there's indication there is clustering. Then I pull these two slabs together --

COMMISSIONER HANSFORD: Sorry, why 5 per cent? I thought it was 3.5 per cent you were referring to.

A. 3.5 per cent is our upper bound. I'm doing hypothesis testing, you compare p value against significance level, that's what I call gold standard. That's the most commonly accepted threshold value for hypothesis testing.

COMMISSIONER HANSFORD: Right.

A. So while it combines the two slabs' data together, it shows no clustering effects. I have four bullets here, I don't want to elaborate on that. Basically people can go there to see how I did it.

Next slide, please. Now I'm focusing on Dr Wells' second point. What if the data are indeed clustered and then you mistreat it as independent? What would happen? Suppose the data indeed is clustered, then if you treat

Page 33

data as independent, then you think you have more data than you actually have.

Okay, that's the wording in red, "you would think you have more data than you actually have". This would underestimate the variance. I will explain all this if you have a hard time to understand, but this would underestimate the variance and thus give you a shorter confidence interval at 95 per cent, and in time the upper bound of the 95 per cent confidence interval would also underestimate the population one.

So my conclusion is, at the bottom, this will lead to lower rates, not higher rates, lower rates of defectives in the sample than in the population. Hence, any results will necessarily be less conservative than should be the case.

So my conclusion is exactly the opposite of Dr Wells' conclusion.

COMMISSIONER HANSFORD: But nevertheless, Prof Yin, you are saying this is academic because you have demonstrated there's no clustering?

A. Yes. I demonstrated in the data --

COMMISSIONER HANSFORD: And therefore this is academic; this is not relevant?

A. It's hard to say because -- go to the previous slide, please. No, the previous one.

Page 34

COMMISSIONER HANSFORD: You said if there's clustering ...

A. One more slide, please. Not this one. One more slide, please.

MR PENNICOTT: The other way. That's it.

A. Yes, exactly. Thank you very much.

So, if you look at NSL, there's a clustering indicated by the hypothesis testing result. 0.007 shows significant result, that's indicating there are clustering effects. In EWL, no statistical evidence for clustering. But when I pull the two slabs' data together, again shows no clustering effect. It depends on how you look at the data. If you want to look at it separately, this could be a relevant point.

COMMISSIONER HANSFORD: Okay.

A. Thank you.

COMMISSIONER HANSFORD: I understand.

A. Next slide, please.

Next slide, please. I'm not trying to amuse everyone here. I am trying to tell you what do I mean by "clustered". Look at the dogs here. Some dogs, they are the same breed, like two dogs in the top corner, five dogs are the same breed. This is what I mean by "clustered data". "Clustered data" means the data are correlated. Imagine if you have all these dogs are from different breeds, that's what I would consider as

Page 35

independent data. This is clustered data. You imagine you replace all those -- you know, the two dogs the same, take one out and put another breed. If all these, say, 20 dogs are all distinct breeds, I call that independent data. I just give you an analogue to understand the meaning of clustered or independent. Then you can imagine if I have the same number of dogs but they are all of distinct breeds, that gives you a lot more information about the dogs.

Next slide, please. So this, graphically, I want to illustrate the points in my previous slides. Look at the top corner on the left. It's a sample size of 100 clustered data. Because the data are clustered, it actually is equivalent to a sample size of say 80 of independent data. But what you have done is you think this 100 clustered data being mistreated as if they are independent. Still sample size 100 but you mistreated 100 clustered data to be 100 independent data, so you look at the arrow pointing down, that's what you -- misconception.

So, with this inflated sample size, it will cause you to have a narrower curve, on the right side, which is in red. So your estimate would have a narrower curve, because you would have lower variance. But the truth is the blue curve. You misconceptually infer that

Page 36

100 clustered data to be 100 independent data, so you think you have more information than you actually have. So, basically, you underestimate the variability. That would cause you have a lower defect rate. You would underestimate the defect rate. Basically, the defect rate, I'm talking about the 95 per cent confidence interval upper bound.

Next, please. So another point about how do you handle missing data, because Dr Wells spent a lot of writings about how to handle missing data.

So using the mean value to impute the missing observation would inflate the effective sample size, because the data are missing, but you said, "I'm not considering they are missing. I'm going to use the rest of the data that are not missing, calculate the mean; I use that mean to impute all those missing observations." Clearly, you enlarge your sample size, because if you throw away those missing data you would have a smaller sample size. Now you impute those missing data, you would include those missing data in your analysis, you would have larger sample size and thus a smaller variance.

There are several problems here. You use the same value to impute all missing data, you do not account for variation in the missing data, because you are assuming

Page 37

that all this missing data actually are equal, but the underlying true data, nobody knows; they could be very different. But you impute it using the same value again and again and again. Basically, you reduce the variability of the whole sample.

The third bullet here: deleting missing data gives valid results under the missing completely at random assumption. What do I mean by "missing at random assumption"? The missing data was caused by the PAUT results not obtainable, not obtainable because of engineering problem, maybe due to angles, smoothness of the surface, ultrasound problem. So it's not because of the data itself. It is because of something very irrelevant to the data itself.

I went to the Hung Hom Station site. I looked at the couplers. You have to make sure the coupler surface is very smooth and shiny, like a shoe, in order to have the PAUT results working. So, basically, all I'm arguing here is that deleting the missing data does not bias your sample and actually you should not impute the missing data using the mean value. Here at the bottom I give you a reference. I hope my counsel has please distributed this: "Reference: 'Three problems with mean imputation'." (Handed).

COMMISSIONER HANSFORD: Thank you.

Page 38

A. So this, basically -- just one print, very simple:

"3 problems with mean imputation."

The bullet points:

"-- Mean imputation reduces the variance of the imputed variables.

-- Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval."

The third bullet is actually quite irrelevant, is less relevant. Basically, it does not preserve the relationship, for example, correlation, but that's not relevant. The first two bullets are the most important points I want to point out.

Next, please. In Dr Wells' paragraph 4.11, he was talking about measurements were only taken if visual inspection is passed.

My understanding, visual inspection is to check whether couplers connected or not. If the coupler is not connected, you don't need to do PAUT. The engagement length is clearly zero. This is not a missing data. You have a valid data point, which is zero. Discarding those samples because PAUT results are unobtainable, for reasons unrelated to the potential outcome -- what is the potential outcome we are trying to obtain? That's engagement length. You throw away

Page 39

those samples due to the reasons not related to the outcome is a valid statistical approach; it does not cause bias.

I give you another example. Suppose you are measuring toxicity level, and some of the patients cannot tolerate the toxicity so they drop out; you cannot take measurements on those subjects. That kind of missing data is not missing at random, because the missingness depends on the outcome. The outcome is toxicity level. If the toxicity level is too high, they all drop out. Then, if you throw away those missing data, that would cause the sample being biased.

So what I say if the missingness is unrelated to the outcome you are trying to measure, here it's engagement length, then throwing away missing data is a perfectly valid approach.

Next, please. There are a lot of discussions about whether you use continuous or discrete. So my understanding, the engagement length is indeed a continuous measurement, no doubt, but it's often critical to make a decision on pass or fail in practice.

I here give you several examples: US FDA makes a decision to approve or not approve a new drug. For blood pressure measurements, you need to decide if the patient has hypertension or not, based on continuous

Page 40

measurement. Hypothesis testing, as I said earlier, often based on 5 per cent significance level, you reject or not reject null hypothesis. And cancer patients are often calculated one-year survival rate, so basically at one year the patient is dead or alive. The bottom one is just my understanding. If, at the beginning, you want to claim, "I'm going to use 33 millimetres' or 28 millimetres' engagement length for all the coupler connections", would any contractor be accepted?

COMMISSIONER HANSFORD: I don't understand that final bullet.

A. Okay. Let me elaborate. But anyway --

COMMISSIONER HANSFORD: Maybe that's not statistics.

A. That's not relevant. Let's move on. I don't want to spend too much time.

COMMISSIONER HANSFORD: I think that's somewhat contentious, that final bullet.

A. Okay. Next slide, please. Then there are also continuous variables or multinomial. Why I choose binomial distribution? First, binomial distribution has minimum assumptions. It's simply "yes" or "no". And binomial distribution can give us an exact method, and I think exact method is very statistical jargon. "Exact method", basically, you do not need to assume you have a huge amount of sample size, and this is the most

Page 41

commonly used variable, for example, patient response or no response, dead or alive. If you choose multinomial distribution, then you would involve more artificial input. For example, you have pass, partially pass, partially fail, fail. And where to choose those cut-off points is getting messy.

Then why not choose continuous measurement? That's one argument I have seen. Actually, if you use continuous measurement itself, that requires some other assumptions. For example, it's often assumed the data follows normal distribution, which actually is hard to establish. The data could be very skewed and asymmetric.

And also you have seen that EWL, the engagement length, there are eight zeros. Those eight zeros are very isolated from all the other measurements. They are clustered around 40 above or below. But those eight zeros would cause -- you need a mixture distribution, not just a continuous distribution. You should treat the eight zeros probably separately.

So I'm talking about even though you can dwell upon continuous measurement to give you better results or whatever, in the end you will encounter many other challenges.

To give you one more example about the continuous

Page 42

data. Continuous data are sensitive to outliers. Why I say that? Suppose you have a very, very big number out there. That will pull the whole mean towards that number. But for binary data, you don't. A very large number, still class it as one.

COMMISSIONER HANSFORD: Sure.

A. That's it.

Next, please. There's also a lot of discussion about the 95 or 90 per cent confidence interval. CS2:2012 was quoted --

CHAIRMAN: Sorry to interrupt. When you are looking at -- just going back to be what you have just done, the binomial, it must depend, would this be correct, on what you are seeking to get statistics about, in the sense that if you are seeking a statistic as to whether it is defective or not, then that's simple enough. It's considered defective or it's not.

A. Yes.

CHAIRMAN: Pass or fail.

A. Yes.

CHAIRMAN: But if you are looking at something like safety --

A. Yes.

CHAIRMAN: -- then you may say that's more of -- it sits on a continuum. In other words, you can't say, "If it's

Page 43

1 millimetre this way it's unsafe", unless you are talking about drugs or something like that, but I'm talking about a big building like this. Do you see the point I'm making?

A. I understand. My research is in clinical trials, so that's why I have a lot of medical examples here. Drugs is serious hypothesis testing. You cannot approve a drug that's not working because that would affect many, many people's lives. It's the same issue here: safety. Drugs is a safety issue too.

CHAIRMAN: But your function was to look at defectiveness --

A. Yes.

CHAIRMAN: -- or lack of defectiveness?

A. Yes.

Next, please.

COMMISSIONER HANSFORD: Sorry, just to follow up on the Chairman's point -- defectiveness or lack of defectiveness based on criteria, pass/fail criteria, that you were given?

A. Yes.

COMMISSIONER HANSFORD: I understand.

A. Actually, I was provided the data already determined defective or non-defective. I was provided an Excel sheet, there's column "Outcome", yes/no/yes/no.

CHAIRMAN: That's right.

Page 44

A. So I have no right to say yes, it's wrong, or no, it's wrong. I didn't do anything like that.

COMMISSIONER HANSFORD: I understand.

CHAIRMAN: So you were told, "This is our determination of what is defective or not, these are our samples, and you must now do a statistical analysis that takes into account the entirety of what we are looking at"?

A. Yes.

COMMISSIONER HANSFORD: So effectively, Prof Yin, you were given -- I don't know if this is the right term, but you were given binomial data? You were given pass or fail data?

A. I was given more than pass or fail data, because you saw the plot I plotted, I was given the engagement length too. So there are multiple columns. There's one final column that tells you pass/fail.

COMMISSIONER HANSFORD: But you just told me somebody else decided whether it was defective or not.

A. Yes, that's what I was provided. All the Excel sheets has more than "yes" or "no". It has a column "Engagement length", it has panel number; it has more information than "yes" or "no".

COMMISSIONER HANSFORD: Okay.

CHAIRMAN: Sorry, I know we are holding you on this a little bit, but -- defective, as you say, is binomial in the

Page 45

sense that you are told this measurement and that measurement for purposes of these statistics are to be considered as amounting to a defective installation; right?

A. Yes.

CHAIRMAN: Which would be different, of course, but we are not talking about statistics here. If we go back to square one, if you turn around to a diligent workman on site and you say, "It's defective if it's 1 millimetre this way", he's probably going to turn to you and say, "This is a rebar which is 4 metres long. It weighs X number of kilograms. It's got to be put into a coupler, and when you are putting in 1,000 of them over a week, believe me, 1 millimetre is not going to make the blindest bit of difference."

So, in other words --

COMMISSIONER HANSFORD: And on top of that he can't measure 1 millimetre.

CHAIRMAN: Yes, "and I can't measure 1 millimetre."

This is not to try to undermine, it's just we are looking at a specific -- you are looking at a dictated set of figures which you are told amounts to defective; right?

A. Yes.

CHAIRMAN: And then you are told, "Please look at all our

Page 46

samples and come up with indications of the degree of defectiveness, based on those samples," which is your job as a statistician?

A. Yes.

CHAIRMAN: Okay.

COMMISSIONER HANSFORD: And the points we are raising are simply around the context of why they may or may not be classified as defective.

A. Yes.

CHAIRMAN: Yes, which is different.

A. Yes.

CHAIRMAN: But then your argument would be, "If that's your case, give me a different set of initial figures."

A. Yes, and all those figures should come from engineer, not from me.

CHAIRMAN: Exactly, so that becomes an engineering question.

A. Yes.

CHAIRMAN: Okay. Good. But of course your figures rest on the shoulders of the engineers then.

A. Yes. Just like if I analyse clinical trial data, I would not challenge the doctor's decision. The doctors tell me, "This is the data I give to you", I shouldn't go back to say, "You are wrong"; okay?

Another whole lot of discussion about confidence interval, whether it's 95 or 90, my understanding here

Page 47

is again, for CS2:2012, they have this manufacturer's test and the purchaser's test. The whole purpose is gatekeeping to reassure the quality of the rebar. This is very different from what we are doing here. Our goal for this whole project is trying to estimate the defect rate of coupler connections. So it's a statistical inference problem, because we want to collect the sample and infer the whole population. We are trying to do inference problem; it's not a quality reassurance problem.

Also, you see, for CS2:2012, you have two layers of test, one is manufacturer's test, then you do on-site purchaser's test. Then you can have a relaxed confidence interval to 90 per cent. But here there's no reassuring some given defect rate has already been estimated by another party.

So my point is, for CS2:2012, you have two layers of test trying to ensure the quality of rebars. Here, we are trying to do a statistical inference. And the bottom bullet is, in hypothesis testing, as I've said again and again, p value would be calculated -- will be -- I use "gold standard" here -- will be compared with gold standard 5 per cent significance level. That's basically trying to control the false positive rate at 5 per cent.

Page 48

CHAIRMAN: Sorry, can you help me -- I don't understand your third paragraph:

"There is no such a layer corresponding to 'manufacturer's test' in our case, and we are not reassuring some given defective rates ..."

In other words, these figures are not meant to support -- sorry, your statistics are not meant to support the correctness of the figures given to you initially; they are just meant to --

A. The whole slide here, I'm just trying to say 95 per cent confidence interval should be used instead of 90 per cent confidence interval.

CHAIRMAN: I see. Okay. So now you are only talking about confidence interval?

A. Yes, I'm only talking about confidence interval right here. And also I want to say the confidence interval should be pre-specified. You cannot look at the data and then change your confidence interval.

COMMISSIONER HANSFORD: Sure.

A. That's basically data -- it's called p hacking or p value hacking. You try to get the results that become significance. So you have to pre-specify your confidence level in advance and you cannot change it after you observe your data.

Next slide, please. This slide is basically talking

Page 49

about the Monte Carlo method. As I said, we use the bootstrap method which is the most popular approach to estimate the variance. We also tried two different versions of the bootstrap, which gives very similar answers. That's it. Thank you very much.
COMMISSIONER HANSFORD: Just on this slide, though, I imagine counsel for some of the parties may raise the point, but my understanding, Dr Wells said that bootstrap and Monte Carlo were addressing different problems.
A. No.
COMMISSIONER HANSFORD: Well, that's what he said.
A. Yes, but I disagree.
COMMISSIONER HANSFORD: You disagree?
A. Yes.
COMMISSIONER HANSFORD: Okay.
A. I think next slide is "Thank you"; right? I just want to make sure.
Examination-in-chief by MR KHAW (continued)
MR KHAW: Just one question. If we can go back to your slide 17, where you talk about the samples in area A; do you remember?
A. Yes.
Q. Then you also mentioned -- and in fact we have also heard evidence regarding the restrictions in area A.

Page 50

A. Yes.
Q. And that is because it's blocked by mass concrete infill between the two slabs, as you have also mentioned here?
A. Yes.
Q. And, as a result, we understand that some panels in area A were actually excluded from the sampling process.
A. Yes.
Q. May I just ask whether you were involved in the decision regarding which panels should be excluded --
A. No.
Q. -- from the sampling process in relation to area A?
A. No. And the third bullet says:
"A list of panels with couplers were provided [to me]."
Then I carry on the random sampling process.
MR KHAW: Thank you. I have no further questions. The lawyers in this room may have some questions for you, and obviously the Chairman and the Commissioner may have further questions.
WITNESS: Okay.
MR KHAW: Thank you.
CHAIRMAN: Who should go?
MR PENNICOTT: Mr Shieh.
CHAIRMAN: Mr Shieh.
Cross-examination by MR SHIEH

Page 51

MR SHIEH: Good morning, Prof Yin. I represent Leighton Contractors (Asia) Ltd and I have a few questions for you.
A. Yes.
Q. You probably know more about statistics than everybody in this room.
MR PENNICOTT: Put together!
MR SHIEH: Together. Could I just set the scene? This is not an academic symposium --
A. I understand.
Q. -- where people sit together and present papers on controversial topics.
A. Yes.
Q. You accept -- I'm not talking about any topics that we have been discussing, but in every respectable academic discipline, and statistics is obviously one of them, there are bound to be areas where people take different views on a legitimately controversial matter; do you accept that?
A. Yes.
Q. And you would accept that unlike primary facts such as which day of the week today is or how many fingers I have, very often, in questions of opinion, you can't insist that there must be a correct answer? As a general proposition, do you accept that?

Page 52

A. What do you mean, you don't have correct answer? There are many things you should have correct answer.
Q. Yes, but there are many things in respectable academic disciplines which are incapable of yielding a correct answer, and that is why we have debates, we have symposiums, we have seminars?
A. No, I wouldn't say so. I think a lot of research can be carried out to determine what is correct or what is wrong, or what is more appropriate and what is less appropriate.
Q. Can I now ask you something about -- let me start it this way. I would first like to engage with you on something which is not your expertise. I would like to engage with you on two topics. One, the binomial approach. Two, acceptance criteria.
First, binomial. To put it in the simplest possible terms, a binomial approach is an exercise whereby every trial or every test would yield two possible outcomes --
A. Yes.
Q. -- pass/fail, yes/no, die/alive?
A. Yes.
Q. Or defective/non-defective?
A. Yes.
Q. So that's binomial, two; right?
A. Yes.

Page 53

Q. Multinomial would mean more than two possible outcomes?
A. Yes.
Q. Maybe pass/partially pass/fail?
A. Yes.
Q. Would academic grades be regarded as an example of multinomial, A/B/C/D/E?
A. A good example, yes.
Q. Without saying pass/fail, because if it's including pass/fail then there's an element of binomial in it, but if you simply say A/B/C/D/E, that would be an example of multinomial?
A. Yes.
Q. I think you have accepted this earlier on, from questions from Mr Chairman, but let me just get it out of the way for the record. Let's imagine, if you are an administrator of a government, you may wish to design a scheme in order to help you decide whether to accept certain applications; right? So you have to devise a scheme for you to tick the box: accept/not accept?
A. Okay.
Q. And in a case like this, you may think that for administration reasons you need binomial, and very often you would accept, would you not, that in terms of helping administration, ease of administration, people would tend to go for binomial approach?

Page 54

A. I want to make sure I understand your question clearly. You want to put me in a hypothetical situation that I'm the administrator or I'm the statistician?
Q. Administrator.
A. I'm an administrator, I'm not a statistician?
Q. You're administrator, yes.
A. Then your question is to design a scheme to accept or not accept?
Q. Yes. If you are an administrator, you have decided upon a scheme -- you want to design a scheme to help you accept certain applications; right?
A. Yes.
Q. And this would be a typical example whereby a binomial approach would be used because it's easy to administer? You have pass/fail, you have criteria that is easy to administer; would you accept that as a general proposition? Or would you think this is outside your area of expertise?
A. No. I think there are a lot of factors need to be put in. I mean, this -- your question is too general to give a specific answer.
Q. Okay. Good. I focus on more specific matters --
A. Okay.
Q. -- concerning the subject matter of this case.
A. Yes.

Page 55

Q. In this case, you were presented with -- well, you are preferring a binomial approach; is that correct?
A. Preferring?
Q. Over multinomial.
A. I think this word is emotional. I wouldn't say I prefer. I don't know what you mean by "preferring". As a statistician, I look at the problem. I take the most appropriate approach.
Q. Let me just put it in another way. For couplers -- because I'm homing in from general to specific, because if you say the questions are too specific, let me just home in -- if they are too general, let me home in on the specific.
A. Okay.
Q. Let's say, on the acceptance criteria that you have been given, 37 millimetres or 40; right? Certain engagement length, certain number of threads exposed.
A. Yes.
Q. If you fail to achieve that, it's regarded as "defective" or "fail"; right?
A. Yes.
Q. Missing it by half a thread would mean a fail; correct?
A. What do you mean?
Q. Missing it by half a thread -- if it's three threads exposed or two and a half threads exposed, then it's

Page 56

a fail; correct?
A. Based on the criteria.
Q. Based on the criteria you're given.
A. Yes.
Q. If by PAUT measurement it's 36.5 millimetres engaged, it would be regarded as a fail, based on the acceptance criteria that have been given to you; correct?
A. Correct.
Q. You have no training yourself as to whether or not a 36.5 millimetre PAUT-measured engagement length could still provide structural support; correct?
A. Yes, I have no training.
Q. You have no training. So it is possible or it may not be the case, you just do not know; correct?
A. I have no expertise in engineering.
Q. Right. If you don't want to answer hypothetical questions, then by all means tell us. If, as a matter of engineering, a 36 millimetre engaged or PAUT-measured embedded length could still provide structural support, then the binomial approach would result in discarding such a rebar because it would be regarded as a failure, worth zero; do you accept that?
A. I want to understand your question clearly. So you said if you have 36.5 millimetres' engagement length --
Q. Yes.

Page 57

A. -- then what?
Q. If, as a matter of science, as a matter of structural engineering, a 36 millimetre engaged rebar still has capacity of load-bearing, for example --
A. I don't know. This is beyond my expertise, as I said.
Q. I know. That's why I'm fairly asking you -- I'm not asking you to accept this to be the case -- I'm just asking you, if this is the case, then adopting a binomial approach would result in discarding a sample which has some load-bearing capacity.
Now, if you don't want to answer hypothetical questions, then just say so and I will move on.
A. Yes, please move on.
Q. Because I would say it's a matter of common sense, but if you -- so you don't want to answer this hypothetical question?
A. I think all these are beyond my expertise. As you said, if it's a 36 millimetre engagement length, based on the criteria it doesn't meet, it's a fail. I follow the two criteria -- actually, it's not I follow -- I was provided the data, already been following the two criteria, the outcome.
Q. But --
CHAIRMAN: Sorry, if I can interrupt here. That's why I emphasised at the beginning: the criteria you are

Page 58

given bears a label, and that label is "defective". What that label means is a matter for the people who have given you the material. That would be right, wouldn't it?
A. Yes.
CHAIRMAN: Exactly. So they have decided what constitutes "defective" and for what purposes it constitutes "defective". Your job is merely to look statistically at the spread of that particular set of figures?
A. Yes.
CHAIRMAN: So it doesn't necessarily relate to load-bearing or to -- it could relate to a time and motion study on workmanship, for example, or it could relate to the degree to which they meet certain required standards. What you are given are figures that relate to a definition called "defective".
A. Yes.
CHAIRMAN: And that's it.
A. Yes.
MR SHIEH: Thank you very much, because I'm trying to eliminate certain matters which could not be drawn from the report.
CHAIRMAN: Yes.
MR SHIEH: Let me follow up, Prof Yin.
A. Yes.

Page 59

Q. So, in your task of providing statistical assistance to the government --
A. Yes.
Q. -- and in providing your helpful report, it was not part of your remit or you did not regard it as part of your responsibility to ask the person giving you the instructions, "Hang on a second, you tell me to accept X millimetres as an acceptance criteria. How about stuff with less than X millimetres' embedded length; isn't it a bit unfair to exclude them altogether?" You didn't raise these questions because these are things you had been given and you just had to proceed on the basis of what you had been given; is that correct?
A. I work in medical statistics extensively. I would not challenge a doctor and say, "Systolic blood pressure below 130, or above 130, you are being classified hypertension, and below, no hypertension", that's a medical decision. I have no right -- you see, 130, why not 128 or 129? How about 131? It's not my expertise.
Q. Thank you. I think we understand very clearly, loud and clear, as to the limits of your task and responsibility.
A. Yes.
Q. Thank you very much for that.
Can I ask you to look at your report.

Page 60

CHAIRMAN: Sorry, Mr Shieh, could I ask you -- I'm looking here, it's now just before 1.00, and if you think -- are we going to finish by 1.00 or just after 1.00, do you think?
MR SHIEH: I'm going to finish at least this topic before 1.00, or even one more topic, depending on how quickly we go.
CHAIRMAN: Thank you. I appreciate that.
MR SHIEH: Prof Yin, in 3.2.2 of your report for the holistic proposal -- for your holistic report -- it's in COI 1 experts -- do you see that, 3.2.2?
A. Yes, I see it.
Q. You say:
"In the design stage of the holistic proposal, I verified the suggestion using a binomial analysis by MTRCL. I considered the binomial analysis appropriate because it uses the minimum number of assumptions. From the statistical perspective, the fewer assumptions one makes, the more desirable is the statistical analysis. More assumptions may introduce more uncertainty as some assumptions cannot be verified easily. If the assumptions made are not entirely true, the conclusion drawn from the statistical analysis may no longer be valid."
Do you see that, 3.2.2?

Page 61

A. Yes.
Q. Then, at 3.2.3, you talked about the question of partially engaged coupler connections; do you see that?
A. Yes.
Q. You say:
"If coupler connections with insufficient engagement can be allowed and taken into account in the design, multinomial analysis may be relevant."
Do you see that?
A. Yes.
Q. Now, there is a big "if" in this sentence; do you see that? This sentence starts off with the word "If"; right?
A. Yes.
Q. "If coupler connections with insufficient engagement can be allowed and taken into account in the design, multinomial analysis may be relevant."
A. Yes.
Q. But you have told us that the instructions that you had been given are that if it's insufficient engagement length, it is to be treated by you as a failure, correct, according to the acceptance standard that you have been given?
A. So let me clarify. There are a lot of discussions, there are many meetings going on at HKU, and I remember

Page 62

all these discussions, multinomial analysis was raised up, and we also put analogy just like A/B/C/D, like student grades. We discussed all these things and in the end we decided binomial approach is the most appropriate and feasible approach.
Q. Because the government indicated that it was giving you instructions on the basis that if certain criteria is met, then it is to be regarded as a fail; if not, if they are passed, then they are regarded to be non-defective?
A. No. At that time, I don't even know the criteria. 37/40, I have no idea. But there is a continuous variable. It's possible you have multinomial, you just put more cut-off points, below 30, below 20. You can have multiple categories.
So I remember all these lengthy discussions, continuous random variable or multinomial or binomial. When we discussed, I have no idea, because all those PAUT results were taken by professional engineers, not my responsibility. So, basically, as a statistician, if you tell me you have a continuous measurement, certainly I would consider you can either classify multiple levels or two levels or use a continuous random variable. This is all possible, under discussion.
Q. Different ways of treating data, let's say multinomial

Page 63

versus binomial versus continuous, have their strengths and weaknesses; correct?
A. Yes.
Q. So you outline certain advantages of binomial; right?
A. Yes.
Q. Fewer assumptions, easier to operate?
A. Yes.
Q. The disadvantages of multinomial or maybe continuous would be it involves more assumptions and maybe more complicated procedure; these are the kinds of -- you would regard them as disadvantages?
A. No. As you said, for every method you are going to choose, there are pros and there are cons. You have to value all the things holistically and take the most appropriate and feasible approach.
Q. Depending on what the ultimate user of the model wants to achieve; correct?
A. I don't know what you mean, the ultimate user -- what do you mean? Who is the ultimate user here?
Q. Who commissioned you in this exercise?
A. I was approached by the government.
Q. So depending on what use the government wants to make of the statistical model, the government would decide ultimately whether or not to go for a binomial model or multinomial model or continuous model?

Page 64

A. No. The government don't decide.
Q. Who decided it?
A. HKU statistical team I led decided. And also we discussed this with MTRC, and then we have lengthy discussion; we decided binomial is the most appropriate and feasible.
You have to -- you see, I use "feasible". Maybe other approaches, you can fantasise about it, you have complicated model, but in the end it's not feasible.
Q. But I'm a bit puzzled, and I'm not going to labour on this point any further because it could well come up to a matter of argument, but -- so you are now telling us that it is HKU's recommendation that binomial model is the most appropriate one to use out of multinomial and continuous?
A. No. I think you -- the decision using binomial approach is discussed among many different stakeholders. We have -- I don't remember all those people's names. I had many, many meetings with many, many different experts. I don't know who they are, where they come from. I know they are either from government or MTRCL. And after lengthy discussion, this is the consensus.
Do you understand?
Q. I will try it one more time and then I will move on. You told us that it is not part of your training to

Page 65

decide upon what particular embedded length would give rise to acceptable measure of structural safety; correct? It's not part of your training; you are not an engineer, correct?
A. I'm not an engineer.
Q. And you won't be able to tell whether or not, by failing a sample with only 36 millimetres' engagement length, you would be discarding a sample which has some load-bearing capacity? It's not part of your training; correct?
A. Yes, I'm not an expert in engineering.
Q. And so you had to rely on someone else to actually tell you, "Look, forget about working out load-bearing capacity. We are telling you that from our perspective, the government's perspective, 37 is acceptable but anything less is not acceptable"? You have to rely on what the government has told you in that regard; correct?
A. I wouldn't say government tells me. I don't know -- the criteria is set up there, I believe that's engineering profession. I don't know whether it's government or MTR. I have no clue who are all the parties I have met.
CHAIRMAN: Let's put it this way: it's those who instruct you who will make that final decision? Well, no, you -- as I understand it, they will come to you with the

Page 66

problem. They will tell you what they are looking for. In this case, it is whether the installation of reinforcing bars into metal couplers were defective or not.
A. Yes.
CHAIRMAN: You will say, "What do you mean by defective?", and they will say, "Well, we've got a set of figures, and it means, if it shows a screw on the outside and certain screw depths on the inside, that's our defective figure", and you will say, "Okay, if you are just looking for defective/non-defective" -- well, no. "If you are looking for defective on a statistical basis, binomial I think will work the best, and is the most feasible because we can make it more complicated but it becomes non-feasible"?
A. Yes, I think it's fair to say that.
CHAIRMAN: So you make that final decision in the light of information given to you?
A. You know, I'm a layman in engineering field. My understanding is if you have a coupler, my understanding, the natural understanding, you've got to insert this thing inside (demonstrating with a pen). It's a valid, sound cap. You cannot hang halfway through. That's my understanding.
COMMISSIONER HANSFORD: That's a matter of engineering.

Page 67

A. No. As a layman -- I'm not an engineer -- you know, I use a pen every day.
COMMISSIONER HANSFORD: With a pen, I would agree with you.
MR SHIEH: A coupler is different from a pen, but I'm not going to argue with you.
A. I know. But look, I can unscrew here, this is what looks like a coupler, I unscrew it, I don't unscrew halfway. Just as a layman, I don't consider this is a good, sound screw-in exercise. I have to screw all the way if I'm going to write with this pen.
COMMISSIONER HANSFORD: This is a matter for --
MR SHIEH: Professor, you understand -- have you given evidence before as an expert witness?
A. No.
Q. You understand that your task is to provide impartial, objective assistance --
A. Yes, I understand.
Q. -- to the Commission and not try to act as an advocate in favour of any particular party?
A. Yes, I understand that. I signed --
Q. The expert declaration?
A. I read all the codes and I understand.
Q. Yes. Thank you very much.
In view of what you have said by way of answer, I may not have anything to add on these two topics.

Page 68

Sorry, Professor.
COMMISSIONER HANSFORD: Can I take it one step further.
Prof Yin, if you're given a set of data, and through discussion a decision is made that the best form of analysis is binomial, does that then mean you must be given the pass/fail criteria?
A. No. All this discussion -- I was provided data --
COMMISSIONER HANSFORD: No, I'm sorry, I'm not saying that you had been. What I'm saying is if you take a situation where you are given some data, and through discussion a decision is made that the best form of analysis of this data would be to use a binomial analysis, it would then be necessary to be given pass/fail criteria. That would be the next step that would be required in order to then be able to carry out that analysis. Am I correct?
A. It depends. If you have a survival study, it's dead or alive. There is no criteria. Just dead or alive --
COMMISSIONER HANSFORD: That's a very easy one, but we are not talking about dead or alive. There's no pulse on a coupler, unfortunately.
A. Yes.
COMMISSIONER HANSFORD: So, therefore, someone has to determine a pass/fail criteria, and what I'm trying to understand is: that then comes after the decision is

Page 69

made, or it would appear, after the decision is made that binomial analysis would be used -- that's what I think you are telling us?
A. Yes, after.
COMMISSIONER HANSFORD: After. So step 1, if I'm understanding this correctly -- tell me if I've got this right -- is receipt of data. Step 2 is discussion about this data and decision about what's the best form of analysis, and the decision was taken, through discussion, that the best form is probably binomial. Actually, let me remove the word "probable" because that's not a statistical word. So the best form of analysis is binomial. Then step 3 is, therefore, we need pass/fail criteria, and that was then given by engineers. Is that correct?
A. No. You said step 1 is receiving the data.
COMMISSIONER HANSFORD: Yes.
A. No. Before we receive the data, we already have lengthy, lengthy discussions.
COMMISSIONER HANSFORD: Okay.
A. Before I see any data --
COMMISSIONER HANSFORD: That's fine.
A. The data was provided at the end, in the end of the whole thing.
COMMISSIONER HANSFORD: Okay. So step 1 is a discussion

Page 70

about what data will ultimately be received.
A. That's fair to say.
COMMISSIONER HANSFORD: Step 2 is a discussion about that data that will ultimately be received and therefore what is the best form of analysis of that.
A. Yes.
COMMISSIONER HANSFORD: And the decision is made the best form is binomial.
A. Yes.
COMMISSIONER HANSFORD: And step 3 is, in order to do that, we need to be given a pass/fail criteria, but that has to come from the experts, who are the engineers, so that's then provided.
A. I think let me clarify a little bit. Actually the binomial came at the very beginning.
COMMISSIONER HANSFORD: Okay.
A. Because --
COMMISSIONER HANSFORD: That's fine, but that doesn't change my point, does it? My point is a decision has been made that binomial is to be used. Then a pass/fail criteria's got to be given, because you can't do -- my understanding is you can't do binomial without there being a pass/fail criteria.
A. Yes. So the first time I see the whole project, the binomial already arise.

Page 71

COMMISSIONER HANSFORD: Okay.
A. It's a natural approach.
COMMISSIONER HANSFORD: I'm happy with that?
A. Defective/non-defective, the most used example in statistical class, in any course in statistics, flip a coin, very simple. So this is the most natural approach. It arises right from the beginning.
If you give me a project like this, the first thing I think is binomial.
COMMISSIONER HANSFORD: But in order to do binomial, a decision then needs to be taken on the criteria for heads or tails, doesn't it, pass or fail, black or white?
A. Okay. The first time I was approached to assist to investigate the whole thing is they talk about the coupler was cut or not cut.
COMMISSIONER HANSFORD: That sounds binomial to me.
A. Yes.
COMMISSIONER HANSFORD: That's different to engagement length.
A. There's a whole lot of allegations going on, because before opening-up we have no idea what is going on. I have no idea. I don't know whether it's cut or not cut or engagement length. There is no information why I was approached first to see this whole problem. So cut

Page 72

or not cut then --
COMMISSIONER HANSFORD: So if the data provided to you is "cut or no cut", it's clearly binomial?
A. Yes.
COMMISSIONER HANSFORD: But if the data that's provided to you is there are different engagement lengths, is it still clearly binomial?
A. As I said, even if you give me a continuous measurement engagement length, I use the blood pressure example, the blood pressure is a continuous measurement. You have to tell the patient, "You have hypertension, I'm going to prescribe you the medicine."
COMMISSIONER HANSFORD: So therefore someone, in that case a medically qualified person, has to give you a hypertension or non-hypertension cut-off point, a criteria?
A. Yes, that cut-off criteria is 130.
COMMISSIONER HANSFORD: Yes, yes, but in the case here, you have to be given the criteria by engineering qualified people?
A. Yes.
COMMISSIONER HANSFORD: And then you can apply that to the data. But what you're telling us is the decision had already been made to use binomial, before you even saw the data, because you thought it was going be "cut or no

Page 73

cut"?
A. Yes, you see, the data -- I have to explain this. I design clinical trials. You design clinical trials, you don't see the data. You eed to plan out the whole problem right at the beginning.
COMMISSIONER HANSFORD: I understand.
A. How many samples you need, what if this happens, then what do you do? The same situation here. I was approached, I don't know exactly what's going on, cut or not cut.
COMMISSIONER HANSFORD: But when you do the clinical data, you are not told "dead or not dead", are you?
A. No, toxicity or non-toxicity, response or no response.
COMMISSIONER HANSFORD: Okay.
A. If you do a phase 2 trial, this patient responds or does not respond, binary data.
COMMISSIONER HANSFORD: Okay.
A. And there are other possibilities but then you have to discuss all the things in advance, before you see the data. The trial needs to be designed before the trial starts collecting the data.
COMMISSIONER HANSFORD: Yes.
A. Like sample size needs to be calculated before I even look at any --
COMMISSIONER HANSFORD: So before collection of any data

Page 74

here, the decision was, whatever data comes, the method of analysis that would be used would be binomial?
A. You know what, I think, as I said earlier, there are a lot of things you need to pre-specify before you observe the data, otherwise you manipulate the whole analysis.
COMMISSIONER HANSFORD: I understand.
A. Like the confidence interval, 95 per cent, 5 per cent is significance level, you need to pre-specify before you even see anything. And once you see the data, you change it, it's very dangerous. You could manipulate the data to do something that you want the data to tell you. As a statistician, that's what we are trained for.
MR SHIEH: Can I follow up on this? Right at the outset or indeed at any stage -- and this is a question of fact, not a question concerning any statistical expertise; I'm asking you as a question of fact -- has anyone at any stage told you that, "Look, the problem is different engagement lengths may give rise to different load-bearing capacities"? If it inserted 40 millimetres, then it's very strong; if it's 37, it's a little bit strong; if it's 35, it's still a bit strong. So there is a continuum of strengths.
Can you give us some advice as to the best way of working out a scheme to calculate the overall strength

Page 75

or any strength reduction to be applied to be an adequate coupling? Has anyone at any stage told you that the strength or the load-bearing capacity could be a continuous -- bear a continuous proportion with the length of the embedded coupler, and asked you to advise on an appropriate model on that basis?
A. There are a lot of discussions about engagement length. For example, if you choose -- I think we talk about 35, 30 -- I don't remember. You know, there are so many discussions, so many possibilities to choose where you want to cut -- not cut, cut-off values on the engagement length. There are a lot of discussions about that.
CHAIRMAN: All right. Can I just ask this. There were a lot of discussions about this, and at the end of the day, what was determined was that two criteria would be used, right, the ones you have already set out, and there would not be a continuum of criteria? In other words, if it's 35, it's okay; if it's 36, it's getting dangerous; if it's 38, oh dear; and if it's 40, everybody run?
COMMISSIONER HANSFORD: Or the other way around.
CHAIRMAN: Or the other way around.
A. You confused me. The other way around.
CHAIRMAN: Do you see the point I mean?
A. Yes.

Page 76

CHAIRMAN: In other words, at the end of the day, you were told, "These are the criteria, binomial criteria, we wish to use to determine the measure of defectiveness. Now, what defectiveness is is for us. What use we put it to is for us. How good it is in the broader world and how useful it is in the broader world is for us. Your job now is to take these two measurements and to work out a set of statistics to show how those measurements, or the level of their pervasiveness in this particular exercise; right?
A. I was not even involved in those criteria. As I said, I was provided with the data, with the column "pass or fail" already. There's one column, it's called "pass or fail"; it's given already.
COMMISSIONER HANSFORD: With respect, I don't think we are talking about that. I think we are talking about in the discussion stage.
You've told us that there was a lengthy discussion stage, with people from government and MTR, and you don't even know who everybody was, and during that stage you weren't provided with data, you weren't provided with sheets that said "defective"/"not defective". You were planning what to do with all the data that would eventually come.
A. Yes, before I saw the data.

Page 77

COMMISSIONER HANSFORD: And that's the bit we're talking about.
A. Okay.
COMMISSIONER HANSFORD: That's the stage we're talking about. So, during this stage, we're trying to understand how the decision was reached that it should be binomial, and Mr Shieh has just asked about whether there was discussion about different engagement lengths and the contribution that given engagement lengths might have to structural integrity, how that could be analysed, whether there was any discussion of that nature.
MR SHIEH: At all. Can I just follow on from Prof Hansford's question, because otherwise the focus of the question could well be lost. So let me ask this question --
COMMISSIONER HANSFORD: Sorry about that, Mr Shieh.
MR SHIEH: No, it's my question the focus of which has been lost --
COMMISSIONER HANSFORD: Okay.
MR SHIEH: -- because just now I was asking you this question and I will repeat it. At the planning stage -- we are not talking about even before you received the data, at the planning stage, when you planned whether to use binomial or whatever nomial or continuous -- in the

Page 78

instructions or the education process that had been given to you in this area, have you been told -- I'm not talking about exact words, my focus is not on the exact words, my focus is on the idea -- has anyone told you, "Look, the behaviour of metal is such that the deeper you embed, the stronger it is, but then there is a gradual fading out, until, if you completely disconnect, then there's no force, but there's" -- let's say, for example, not precise words, don't catch me on precise words -- so let's say if it's 40, then it's very strong; if it's 35, it's less strong but still quite strong; 32, a little strong but still okay -- so can you design a best scheme to work it all out, would you do binomial or multinomial or continuous? The same thing of this nature taken place. Don't tell me there's been lots of discussions, I know about that already, don't give me that answer. Has anyone discussed with you the behaviour of screws according to different embedded length? I'm talking about that sort of specificity. That is the focus.
I'm not asking you -- I repeat one more time; I'm sure you are an intelligent man -- I'm not asking you whether there have been discussions. We know there have been many discussions so don't open your answer by saying, "We discussed many times". Focus: has anyone

Page 79

discussed with you or given you instruction that "There is -- the behaviour of metal is that it's all gradually fading out and not an abrupt cutting off at a certain length"? Has anyone told you that and asked you to design?
A. That's common sense. I don't even need people to tell me. I understand. I have a physics major, masters' degree in physics. I understand. It's a gradual process. I don't need people to instruct me or to tell me. That's a common-sense thing.
Q. So nobody told you to devise a scheme to take into account the behaviour of embedded threads depending on how many millimetres have been embedded?
A. You see, if you enter that zone, talking about engagement or residual, what kind of association you are talking about? It's linear, non-linear? It's getting into a very complicated discussion.
COMMISSIONER HANSFORD: That would be for the engineers.
A. Exactly. That's why --
COMMISSIONER HANSFORD: No, no, no, sorry. What they do with the data or the analysed data would then be for the engineers.
A. I know, yes.
MR SHIEH: But what I'm asking is: has anyone actually told about that kind of information and asked you to design

Page 80

a scheme accordingly? I know it's complicated. Maybe it's because it's so complicated nobody has told you, but I'm asking a question of fact.
A. No. As I said, I don't even need people to tell me. I understand. This is common sense. Even as statistician, if you give me engagement length, I would automatically think, okay, first, whether you can use it as a continuous variable or if you want to classify into different levels, just as you described, it's a common-sense thing.
Q. But nobody actually gave you the data and some calculation such as if it's 40 then it's X load-bearing strength; if it's 38 then it's a little bit less -- nobody gave you that kind of information; correct? It's a question of fact. Did anyone give you that kind of information and ask you to design a scheme or a plan a model?
A. It was -- this is September; right? It was last year. I don't remember who gave me or who didn't give me. I cannot recall.
Q. Thank you very much. Can I ask you one last question before we break for lunch?
A. Yes.
Q. It's a common-sense question. I hope I can get a simple answer. You teach at a university; correct?

Page 81

A. Yes.
Q. If you set a pass mark at 80 per cent, you are going to get more failures than if you set your pass mark at 50 per cent; correct?
A. Yes.
Q. So whether someone fails or passes very often depends on how high you set the pass mark; correct?
A. Yes.
MR SHIEH: Thank you very much.
CHAIRMAN: All right. Good. Thank you very much, Professor. We are going to adjourn now for lunch, but you will be required to return to give your evidence, to finish your evidence, this afternoon. My apologies for that.
You should be told, all witnesses are told, whether expert witnesses or not, that you mustn't discuss your evidence over lunchtime with anybody.
WITNESS: Yes.
CHAIRMAN: Obviously you can sit with people, have lunch, talk about whatever you like, but you mustn't start debating what you said or didn't say, what you should say, or anything like that. Okay?
WITNESS: Okay, yes.
CHAIRMAN: Mr Pennicott?
MR PENNICOTT: I don't know how Mr Shieh is getting on but

Page 82

I was going to say 2.30. I appreciate it's 1.20 already but I think we should start at 2.30.
CHAIRMAN: All right. 2.30. Thank you very much.
(1.23 pm)
(The luncheon adjournment)
(2.40 pm)
MR SHIEH: Good afternoon, Mr Chairman and Mr Commissioner.
Prof Yin, a few topics, hopefully to get through you reasonably quickly.
First of all, missing values. You remember the topic of missing values; right?
A. Yes.
Q. In fact, you and Dr Wells have both given your opinion --
A. Yes.
Q. -- on how to treat an invalid sample, whether or not you discard it or whether you take it into account but assign a value to it. You have given your views on it.
A. Yes.
Q. I think what I can do is just to put to you what Dr Wells has said and invite your comment.
Dr Wells' view is that by discarding a sample which visually is connected to a coupler, because you have difficulty in measuring the embedded length, you are discarding a sample which potentially could be

Page 83

a non-defective one and therefore it is biased. Therefore, to address that bias, what you should do is to assign the mean value of the other known samples to it.
First of all, this is Dr Wells' proposition. Which part of it do you agree with or not agree with?
A. There is no part. I just don't agree, the whole thing.
Q. But do you accept that under the model that you have designed, a sample will be discarded if it is visually regarded to be connected but which cannot be measured by PAUT; that's correct, yes?
Let's start again. If it's visually unconnected, if by visual inspection it is not connected, it would be put in the "fail" category; correct?
A. Yes.
Q. If it is visually connected, you then proceed to examine the number of exposed threads and to conduct measurement by PAUT --
A. Yes.
Q. -- on the embedded length; correct?
A. Yes.
Q. And if you encounter difficulty in conducting PAUT measurement, you would regard that as an invalid specimen or sample; correct?
A. Yes.

Page 84

Q. So that would not be taken into account in working out defective or non-defective rate; correct?
A. Yes.
Q. It would not count as one of -- it would not be counted as part of the denominator; correct?
A. Yes, correct.
Q. Dr Wells' opposition is that, in doing so, you are disregarding a specimen which potentially could pass, because it is connected. It could very well be 40 millimetres embedded, but you don't know, so you discard it completely, and he says you are disregarding a specimen which potentially could pass. Do you accept that?
A. It's also potentially a fail, potentially you don't know. It could be a pass or could be a fail. This is exactly what I said. What I mean is you are discarding -- you cannot obtain PAUT result, not because underlying outcome pass or fail of the engagement length. You are discarding those samples because PAUT results cannot be obtained, for reasons that's irrelevant to the outcome.
For example, I was told by the engineer, "You need to make the surface of the coupler smooth, very smooth, shining, and the angle of the device has to be properly aligned." It's tedious work to do this measurement and

Page 85

you know the concrete makes those couplers dirty, it's hard to clean that. So this reason is not related to the outcome.

So my statement is discarding those missing values is valid approach. Imputing those missing values using the mean value is invalid approach. I showed you this document, one-page document. There are three points that telling you it's invalid approach.

Q. Prof Yin, two points to follow up. One, your proposition is it is acceptable to discard results which are not obtainable for reasons unrelated to the potential outcome; right?

A. Yes.

Q. But what I'm suggesting to you is in the situation that we are concerned with, in the model of testing, the disregarding is for a reason related to the potential outcome because, by the time you decide to discard an unreadable sample, it has already passed the visual examination, so it is something which is not a clear fail; it is something which has the potential of passing. So it is disregarding a sample which is on the potential pass side of the situation.

A. It's also a potential fail. You don't know. This is my point. This is a very well-known statistical fact: you should not impute mean value to the missing data. This

Page 86

is a very well-known statistical fact.

Q. But it is not a clear fail; right? If it's a complete disconnect, then it would simply be dumped into the "fail" pile; correct?

A. Yes.

Q. So it is not a clear fail, it has the potential of passing, and you are disregarding the probability of it passing?

A. It's also a potential fail. As I said, we have no idea.

Let me give you an example. When -- you cannot discard missing data. I will give you a very simple example. Suppose you are measuring toxicity level of a patient, and the patient is treated by the drug, and some patients, their toxicity level is just so high, they cannot tolerate anymore; they drop out of the study, and you cannot get measurement of their toxicity level. So, by discarding those samples, you actually underestimate the toxicity level because those potentially missing data actually have very high toxicity level.

So that's what I mean. If you discard samples for reasons related to the outcome, you cannot discard them. If you discard samples missing because of reasons unrelated to the potential outcome, it's a perfectly valid statistical approach to discard them, without

Page 87

causing any bias.

Q. Thank you, Professor. I think Dr Wells disagrees with you on the basis that it is a discarding on a ground related to the potential outcome, but you have made your position clear.

My next point is the handout you produced, the document called "3 problems with mean imputation". I just came across this this morning after you handed it out. Within the limited time available -- the author referred to a previous article when he showed how to use SAS to perform mean imputation. Do you see that first sentence?

A. Yes, I see.

Q. So mean imputation is something of a known procedure in statistics, according to this; correct?

A. For simplicity.

Q. Yes. But he says there are three problems and then he tried to explain what he regards to be the problems; correct?

A. Yes. Correct.

Q. So it's an accepted procedure in statistics. This author regards there to be some problems, but it does not mean that it is in all cases inappropriate to apply it for analysis purposes. It doesn't say so.

A. It listed three obvious problems with this approach, and

Page 88

these three problems, in my view, statistically speaking, are serious problems. I would rather discard those missing data without causing any statistical problem to the analysis.

If you impute like this, you are actually causing bias. As he said here, first, you reduce the variance, which I shouldn't. Second, you shrink the standard error, so that invalidates your calculation of the confidence interval, which is our key point here, confidence interval. That's the second point right there.

Q. By applying a mean value to the missing specimen, it actually provides, in crude layman terms, a best guess, a best estimate, as to the likely value in that sample; do you accept that?

In other words, rather than taking that out of the denominator and enumerator, you put that in as part of a denominator, and for the value you just attribute a best guess, based on the mean of the other known values?

A. Yes. Just reflect what you said: it's the best guess. It's not your data. You are imputing something that's not there. That's your guess. And that guess caused a lot of problems listed here.

The mean value imputation is used sometimes for

Page 89

simplicity. Statisticians sometimes use invalid method for simplicity sometimes, because, if you don't have the best solution, you can come up with an approximate solution, rather than no solution.

But here we have a solution, by discarding those samples, and that's a perfect solution.

Q. Right. I think we have put and understood each other's position and can I just move on to the next topic --

A. Sure.

Q. -- and that is the question about randomness.

A. Okay.

Q. Dr Wells' point -- you know his point concerning at the number of panels and the proportion between the panels with capping beams details and without --

A. Yes.

Q. -- comparing with the proportion between the ultimate specimens -- you know the point he is making?

A. I know, yes.

Q. Therefore I call that a randomness point. His point is that upon drawing the specimens and seeing the ultimate outcome, and upon seeing this disproportionality or disparity in the proportion, one ought to pause and reflect whether anything has gone wrong or one ought then to revisit the procedure one has taken? That is the limit of what he was proposing?

Page 90

A. Yes.

Q. It is something which should raise eyebrows and cause you to pause and think and reflect; that is the point he is making?

A. Yes.

Q. As I understand you to be saying in the slides that you have produced, you said sometimes, if you go deeper --

A. Yes.

Q. -- although every panel, if chosen, you examine three couplers --

A. Yes.

Q. -- but if it's buried a few layers down, are you suggesting that, under your methodology, you don't just examine the chosen bottom layer, you also look at couplers in the layers above the chosen bottom layer; is that what you are saying?

A. Yes, because in order to reach the third layer, you have to open up the first layer. That's the data already exposed. You shouldn't throw away valuable data.

Q. So you are saying, in the methodology, if for example upon phase 2 -- that's phase 2; phase 2 is to decide the layer; correct?

A. Yes.

Q. If in phase 2 it was decided that the drawing of lots or the random process, it comes out "the third layer" --

Page 91

A. Yes.

Q. -- you open up the first layer, you see three couplers already --

A. Yes.

Q. -- and you take those three couplers into account?

A. Yes.

Q. So that's why it adds up to more than 84; that's what you say?

A. Exactly.

Q. You also make the point, I think -- correct me if I am wrong -- that different panels have different lengths so they may contain different number of couplers --

A. Yes.

Q. -- is that your additional point?

A. Yes.

Q. So you are suggesting, are you not, that the number, the original number, of 237, should not -- how should I put it -- one should not place weight on this initial number of 237 in assessing the proportion of the ultimate samples; are you suggesting that?

A. What do you mean, "put weight"? Can you be more specific?

Q. Can I show you the way Dr Wells has put it.

A. Yes.

Q. It is Dr Wells' report at paragraph 4.2, internal

Page 92

page 4.

At paragraph 4.2, he said:

"Sampling is a difficult subject ..."

And paragraph 4.3, he said -- he quotes from a document called the "capping beam document". This is a document supplied, I think, by the MTR to explain in greater detail how the MTR -- the details of the sampling conducted by the MTR on those panels in which capping beams are present; right? And MTR explained the formula that was adopted to work out the rate of defects and also the strength reduction factor; right?

A. Yes.

Q. You are aware of that document?

A. I'm not aware of this document.

Q. But anyway, the document actually sets out what is called the Formula, capital F. You are aware of a concept called the Formula, which is a formula used to calculate the strength reduction factor for those panels with capping beam details?

A. I'm not aware of capital F. I don't know what you are talking about.

Q. Anyway, I will take you to that.

A. I'm getting lost. Okay.

Q. The document entitled "D-walls/platform slab connections via capping beams".

Page 93

A. Can I see this document, please?
Q. Yes. It is in the opening-up bundle at 9805.
A. Okay.
Q. You can see:
"From the construction record (out of a total of 237 D-wall panel), 175 ... are without capping beam details (type a) and 62 ... are with capping beam details ..."
Do you see that?
A. Yes, at the bottom.
Q. This is the proportion relied on by Dr Wells to begin with.
A. Okay.
Q. So he says, under the original drawings, 237, and there's a certain proportion of 175 versus 62; yes?
A. Yes.
Q. Dr Wells went on to say -- if you look at the results at 4.4 of the report --
A. Yes.
Q. -- he said:
"A random sample of size 90 found 83 type A ... and 7 type B ..."
Do you see that?
A. Yes.
Q. He says there's a certain proportion between the actual specimens drawn?

Page 94

A. Yes.
Q. And you told me that you understood Dr Wells' point to be that there is a kind of disparity between the proportions between the type A and type B details in the drawings; right?
A. Yes.
Q. Within the 237, there is 175 versus 62, but in the actual specimens drawn it's 83 versus 7.
You understand his point?
A. But his point --
Q. You may not agree with his point. I'm just asking you whether you understand this to be point he's making.
A. You're right --
Q. He is comparing the proportion?
A. Yes. I understand.
Q. What you're trying to say is that you shouldn't really rely that much on the proportion of 175 versus 62, because that may have no bearing on the number of specimens drawn. Is that the point you are making?
A. I have several points. First, you use 175:62 as a population, the number of panels. First, some of these panels are not in the random draw. They are even not in the random sample process, because of two reasons. Either a through bar was used, or there was mass concrete made those couplers inaccessible, and

Page 95

those panels were excluded before we started the random draw. That's one point. So that's numbers should not be used.
Second --
Q. Should not be used?
A. Those numbers -- used --
Q. You mean 175 and 62, those numbers should not be used by Dr Wells for comparison?
A. You should exclude those panels that have already been excluded from the beginning.
And a second point, I think his calculation is invalid because he used 90. That basically 83, type A; 7, type B. These numbers are not original number of samples we drew from the population, because those numbers are the valid PAUT results by removing those missing data. So what he would do, in a more craft (?) way is use 102 as the sample size, and the 11 as the number of samples with capping beam, because that's the sample we took from the job, before doing any missing data removal.
Q. Yes. Focusing on your first reason -- in other words, the 237 -- you know the 237 figure?
A. Yes.
Q. Which breaks down into 175 and 62.
A. Yes.

Page 96

Q. That figure, you say, should not be relied upon because that figure does not reflect the actual population that is available for picking, because some of those panels have been excluded; correct?
A. Yes.
Q. Can I then ask you to look at the opening-up bundle, page 9805.
This document then sets out to explain the "Estimation of overall proportion of failed couplers connections via capping beams at EWL", and there followed a series of calculations. Take a look at that.
You have seen this calculation before; right?
A. Yes.
Q. I think you spoke to this calculation in your report?
A. Yes.
Q. It has been referred to as the Formula, capital F, but you may not be aware of that terminology, but leave that to one side. You are aware of this formula; right?
A. Where is the capital F?
Q. It's not referred to in this document, that's why I might have confused you.
A. I see.
Q. But you are familiar with this process of calculation, Prof Yin?
A. Yes.

Page 97

Q. It ultimately gets to page 9807.
A. Yes.
Q. Which is the strength reduction factor of 68.29 per cent; yes?
A. Yes.
Q. So, basically, these few pages work out the derivation, how the final strength reduction factor of 68.29 per cent was reached; yes?
A. Yes.
Q. Now, in this process, the starting point, at 9805, under the heading "Estimation of overall proportion of failed couplers connections via capping beams at EWL", is actually the ratio that we worked out was in the 237, at the bottom of 9805.
A. Yes.
Q. You start off -- now, I don't pretend to understand the calculation, I only know what numbers have been taken into account -- the starting point of this calculation of working out 68.29 per cent is first of all you look at 237 being the total population of drawings of panels, and then 175:62. You then work out two quantities, called Qa and Qb. Qa and Qb then gets fed into various formulae, over the next page, with p and then various things. Then it goes on and on.
If you look at the bottom of 9005, Qa is 0.7384;

Page 98

yes? Qb is 0.2616; correct? Do you see that?
A. Yes.
Q. These are derived from the very number of 237 and the two numbers of 175 and 62; yes?
So 0.2616 actually features subsequently, in the middle of 9806?
A. Yes.
Q. Immediately above "Result with 95 per cent confidence interval". So that 0.2616 was utilised.
Without actually understanding the magic of all these formulae, the short point I want to make is the very formula that was used to derive the 68.29 per cent itself started with the 237 and the proportion within that 237, which you just told us should not be relied upon.
A. Yes. That's what I said.
So let me explain. If you look at my report, page 20, you can start from page 19 and read on to page 20. So this is the calculation at EWL. So paragraph 4.2.3.
Q. Yes.
A. So we had the formula at the bottom, and that moves on to the next page. You can see the pB1-hat equals to 2 over 7, pB2-hat equals 2 over 11.
Q. Yes.

Page 99

A. So we've got pB-hat equals to 41.56.
Q. Yes.
A. Then we move on to 4.2.5.
Q. Yes.
A. Try to calculate the variance of pB-hat. You don't need to worry out how that formula, where that comes from. Through some algebraic manipulation, you get this variance, and then you plug in those numbers, pB1-hat and pB2-hat, you will get this number of 0.0264.
You move on to the next equation. You see that 0.6829. There's no involvement of Qa or Qb which you mentioned in this document, OU9805. This 237 is never involved in my calculation.
Q. Yes, but if you look at that chunk at 4.2.5:
"Using the delta method and after some algebraic manipulation, the variance of pB is given by ..."
A. Yes.
Q. Is that number of 237 and the components not buried in there somewhere? Because otherwise why, in the earlier document I showed you, which shows exactly the process whereby the 0.6829 is worked out, it actually started off with 237? Why is the purpose of doing all that?
A. That's exactly the point I'm trying to make. I did not use Qa and Qb. If you look at my derivation, step by step, there is no Qa and Qb involved, and this document

Page 100

I am not aware of, I told you already, I have no idea about this F, Formula or whatever. So based on this thing, I think they are trying to do a different calculation from a different perspective. I simply want to tell you this. You could reach the same conclusion through different angles, because there is not just one perfect way to get the right answer. I cannot tell you in detail how this whole thing being worked out, because the symbol used in this OU9806 is not a mathematical symbol.
For example, Qa, capital Q, small a, we don't use this kind of notation. It makes a mathematician very hard to read. This uses two letters. If you look at my documents, I don't use this kind of notation in derivation. Like pB1, what is that? You see p-hat B. He uses all these double letters trying to -- or even sometimes three letters, to denote one symbol, which makes a statistician very hard to understand what he was trying to do.
So I don't want to get into detail about how these calculations. I want to inform you that my derivation does not involve Qa or Qb -- actually I don't even know where this 237 came from. Until you pointed out this document, I had no idea where this number comes from. I just took from Wells' calculation. He mentioned

Page 101

something, probably this capping beam document, but I don't have this document. And then based on his number I tried to work out his reasoning, and I found there are some flaws --

Q. You mean you were not given the very document upon which the MTR purported to justify its calculation of the strength reduction factor?

A. I simply say I don't have these capping beam documents you mentioned here. Where is that? Dr Wells -- where? Yes, here. You see, this 4.3, UO9805, this whole thing, I don't know this document, and frankly speaking, there are so many documents, I just have no time to go through them. I have no time. I have to do teaching, I have to do research, I have to publish papers. No time to dig into so many things. This is my point.

Q. Are you telling us you were unable to understand why, in the capping beam document, they actually started off with --

A. I understand. Until you pointed out to me now, I cannot understand immediately. I have to read through to derive it. Then I can verify whether it's correct or not. I need time. But it's not I'm not capable of understanding this. I just don't have time right now.

Q. So, at the moment, you can't assist us as to why that document had started off with the 237 figure and the

Page 102

proportion, and after having gone through a certain process came out to exactly the number that you worked out?

A. Yes, I did not use those numbers. That's the number I came up. And I don't know who did all these things, eventually they used Qa/Qb, they got this number. What do you want me to say?

Q. It's just that if you have not seen it before, then I can't press you on that. I'm just putting down a marker that the very author of the holistic report, who worked out the 68.29 percentage, actually started off with that number.

A. Yes, but if you look at my report, I had a very much simpler approach to get that number, and I only have half a page of derivation. This one drags on two pages.

Q. Are you sure, in none of the hidden methods that you had -- because in working out a sum, you sometimes use formula or assumptions or some basic underlying methodology which is not written out in numerical form. You're sure that in those underlying formulae or methods that you use, it has not involved utilising 237 and the proportion of 175 versus 62?

A. Yes, I'm sure. I can go through with you again. You can point out which number it is that Qa and Qb from page 19. I laid out line by line. You can point out

Page 103

which number corresponding to Qa and Qb.

Q. No, I'm just saying, in 4.2.5 and 4 -- yes, under 4.2.5, you had a number of references to Delta method and algebraic manipulation, and then further down, normal approximation, and then upper bound?

A. Yes.

Q. Amidst all this, you are sure that you have not utilised that proportion?

A. No, it's very clear. If you look at what is the pB1-hat and pB2-hat, it's given right above. PB1-hat is 2 over 7, PB2-hat is 2 over 11, and that variance is simply plug in these two numbers. And you see there's number 7, there's number 11, that's a sample on the capping beam side. Why is the capping beam side and the other is slab side? That's just a sampling size. Because you have a 7, therefore missing value, you should throw them away and -- look, it's very simple. I just don't see where is the Qa and Qb, where those numbers -- in my formula I don't have those.

Q. Anyway, you have told us that you have only seen the capping beam document for the first time today and within the short time available -- I'm not asking you to do it now, you've just told us that you can't immediately work out why that document, with that number, came up to that -- I'm not going to press you on

Page 104

that because this is not a statistical quiz of wanting you to do a derivation.

A. Oh, you know what, I can tell you now -- just look at OU9806, the bottom, "Result with 95 per cent confidence interval". You first calculate the "Variance (p-hat)", and that "Variance (p-hat)", you keep going down and then you see "Variance (pb-hat)". You see, that "Variance (pb-hat)" I believe is my "variance of pB-hat" in 4.2.5. You see, that's exactly the same kind of formula we use. We've got the same number, 0.0264. I got 0.0264.

So, basically, my understanding is all the calculation above, basically, from this 237 and keep going down until "Variance (p-hat)", those are something else, not pB. My understanding here, "pB" means the probability of failure at capping beam. That's my understanding, that's where the "pB" comes from, "pB-hat"; that's basically my derivation, "pB-hat". So all the derivations above, above with the "Variance (p-hat B)" is something else. So from this point, at the bottom, until the rest, if you look at that it does not involve Qa and Qb at all. It is the same derivation.

Q. But Qa and Qb was utilised to derive a certain other value which was then fed into the ultimate equation;

Page 105

correct? Otherwise, there's no point going through all of that.
A. But you know, as you said, I don't exactly know all these derivations at this moment. I can verify for you later.
Q. Anyway, let's not --
A. This becomes a --
Q. Purely as a matter of intuition, you say that figure should not be relied upon, but in the very formula the MTRC gave us, that was the starting point.
Now, you've shown me your calculation which did not appear to utilise that number, and I was wondering whether, as part of the process in a certain technique that you have used, maybe you have utilised it, but it's unfair for me to put it to you immediately now. Perhaps we will move on. If anything --
A. No, okay. I just spotted one thing. Come back to OU9806.
Q. Yes.
A. If you look in the middle, the two lines above "Result with 95 per cent confidence interval", two lines above, you see:
"From the result of investigation, p-hat b1 equals 2/7, and p-hat b2 equals 2/11".
Those are the numbers given. Then you can plug

Page 106

those numbers into the bottom equation, "Variance (p-hat b)". With those two numbers, you can immediately solve this whole thing. There's no Qa involved and no Qb involved. You can solve this "Variance (p-hat b)" immediately, you get 0.0264, and that is what I have done in my report.
Q. Anyway, I'm not going to argue with you, but if you look immediately below the heading "Result with 95 per cent confidence interval", you do have "Variant" and then "Qa2" and "Variant (pa)". So Qa2 does feature and Qb2 does feature.
A. That, in the formula you try to calculate the variance p-hat, not variance p-hat B. In my understanding, pB is defect rate for the capping beam, that variance of p-hat -- I don't know the definition of p-hat. I think basically p-hat is a combination of -- okay, let me tell you why. P-hat is a combination of the capping beam together with non-capping beam pooled defect rate. If you want to calculate a total defective rate, then you would need some formulas like the proportion, but if you only focus on capping beam alone, then his derivation and my derivation are exactly the same.
So, basically, he was trying to derive something different from what I'm doing, and he also included what I'm doing, what I did in my report. So actually he did

Page 107

two things. I did one thing.
COMMISSIONER HANSFORD: So that I can just understanding this -- and who is the "he" we're referring to here?
A. I don't know. Who wrote ...
COMMISSIONER HANSFORD: It's appendix 2.
MR SHIEH: This document, appendix B to a response given by the MTR.
COMMISSIONER HANSFORD: Yes.
MR SHIEH: We can trace the origin of it, because this is a response to -- because what happened was the holistic report was not very informative as to how the 60-odd per cent strength reduction percentage was given, and we asked for some information, and we were given, as part of the information given to us, this capping beam document.
I can supplement information. My learned junior is checking.
COMMISSIONER HANSFORD: No, I mean, it's also -- and maybe it's a duplication; I'm not sure -- it's also appendix 2 of the MTR's report on statistical analysis.
MR SHIEH: Yes.
COMMISSIONER HANSFORD: That's where it is, appendix II.
MR SHIEH: Yes.
COMMISSIONER HANSFORD: I just wondered who the author would be.

Page 108

MR SHIEH: So it is an MTR document, and it may be by "he" he is referring to the notional author within MTR.
COMMISSIONER HANSFORD: Yes. Maybe my question was a bit rhetorical, but yes.
MR SHIEH: Can I then move on to the next point, which is a reasonably short one, hopefully. You know the point about clustering which Dr Wells mentioned?
A. Yes.
Q. The point is again a rather simple and broad ones. If problems about defects, like inadequate embedded length, are attributable at least in part to poor workmanship, then would it be likely, more likely, that poor workmanship would tend to occur in clusters, in the sense that if there is a worker who is bad at his workmanship, that would tend to permeate the cluster of rebars that he's responsible for in the same locality, in the vicinity of each other. So, if you pick three adjacent couplers, and if the problem of bad coupling is bad workmanship, you pick one coupler with bad workmanship, it would necessarily mean the neighbouring ones are more likely to be badly connected. Do you see the point?
A. Yes, I understand your point.
Q. So that is the reason why Dr Wells says if the reason for bad coupling or inadequate connection is

Page 109

attributable to bad workmanship, then this could be a reason for lack of independence. Do you understand the point?
A. Yes, I understand.
Q. So he's saying, if that is so, then the sampling methodology would lack independence?
A. Yes.
Q. Do you agree with that?
A. But that's a big "if". He has an "if". Can he verify this? Actually, I tried to verify what he was trying to say. He didn't verify. In his report, he posed a lot of assumptions, back-calculating, but I was trying to verify what he was trying to do, and in his permutation test -- this morning, I gave the presentation -- I had this permutation test result. Can we put that up on the screen? My synopsis from earlier this morning? I forgot which page. It's a permutation test with p values.
Yes, right here. That's slide 24. So, actually, I appreciate Dr Wells -- all this enquiry about the statistical methods we have been using and we have been working on, because he actually posed a lot of questions, made me think harder and broader. So he talked about this clustering and then I really get into perform this statistical analysis, hypothesis testing,

Page 110

and if you look at the result, EWL, no evidence of clustering. NSL, there is statistical evidence of clustering. Pool them together, no evidence of clustering. This is the statistical analysis I am presenting to you.
COMMISSIONER HANSFORD: Sorry, I don't understand that. I understand that -- well, I think I understand that this shows EWL has no evidence of clustering and that NSL has evidence of clustering.
A. Yes.
COMMISSIONER HANSFORD: The bit I don't understand is when you pool them together, the conclusion is there's no evidence of clustering. If there's been clustering in NSL; there's been clustering.
A. You are right, but if you think about this, the data being pooled together, the clustering effect in NSL, after pooling them together, probably being diluted and the statistical evidence becomes weaker. So you see the p value actually lies right between these two p values. So p values give you the stress of the statistical evidence, how strong the clustering is. There's evidence of clustering or no evidence of clustering? When you pool them --
COMMISSIONER HANSFORD: I'm thinking of "clustering" in a common-sense term and you are thinking of "clustering"

Page 111

in a statistical term, and maybe they are different. Common sense says to me if there's clustering in NSL, there's clustering in the job somewhere.
A. Yes, you are right. It's clustering in NSL, but then you put NSL in a larger pool and those clustering effects become diluted.
COMMISSIONER HANSFORD: So there's still clustering but the clustering effect gets diluted?
A. The evidence of clustering being diluted by other independent data.
COMMISSIONER HANSFORD: But we don't need more evidence because we've already got evidence of clustering.
A. But when you pool the data together, it's already -- all the information being pooled together to test whether there is clustering or not.
COMMISSIONER HANSFORD: All right.
MR SHIEH: Anyway --
COMMISSIONER HANSFORD: That's why you say, "Permute the data, ie destroy the clustering"?
A. Exactly.
COMMISSIONER HANSFORD: You don't mean "destroy", you mean "dilute"?
A. No, I mean "destroy".
You want to really learn permutation test, I can explain more.

Page 112

COMMISSIONER HANSFORD: No, I will keep that for another day, when I have time to attend a statistics lecture.
A. But I listed these bullets, I think Dr Wells is welcome to follow my procedure to do the similar analysis and to see what's his conclusion.
I listed -- one thing I want to point out: Dr Wells' reports, both reports, don't give much detail about how he gets those numbers, where those table numbers are. It's very difficult for me to verify. So here he can just follow what I did and try to verify whether it's correct or wrong.
MR SHIEH: Prof Yin, can I move on to deal with area A.
A. Yes.
Q. You accept that the strength reduction rate for area A was worked out by a process of extrapolation, because area A itself has no -- yielded no specimen; correct? It extrapolated from data obtained from other areas; correct?
A. You can call it extrapolation, but on the other hand, because area A was involved in the random sample, two panels from area A was inside the pool, but we did not choose them by the random process. So you can either say it's extrapolation, or I don't even say extrapolation. I would say, okay, first area A and HKC have similar configurations. This is based on my

Page 113

understanding.
Q. You were told so by MTR?
A. Yes, engineering.
Q. You were told by the engineers?
A. Yes, engineer said we have similar configurations, same contractors, probably similar workmanship, and since area A already inside the pool, even though no sample were drawn, but it's inside the sample pool. So it may not be extrapolated. It's just an estimate for area A and HKC together.
Q. Yes. Now, can I ask you to look again at the capping beam document that was produced by the MTR. Again, opening-up bundle, page 9810.
A. Yes.
Q. That's the result of the measurement for the EWL slab. Do you see that gave rise to the famous two defects result; do you see that?
A. The bottom formula?
Q. Yes, two defects. Then the total sample for analysis -- I think from the capping beam side there are 11. For the slab side, there are seven; do you see that?
A. Yes.
Q. I think it's the same 11 that's picked, but then for the capping beam side, all are valid, for the slab side four are invalid so you get seven; that's correct, right?

Page 114

A. Yes.
Q. Physically, you talk about the same 11 chunks -- you are talking about the same 11 spots, but then on one side all 11 are valid, on the other side, four are invalid, and therefore discarded; correct?
A. Yes.
COMMISSIONER HANSFORD: Sorry, Mr Shieh, is that the slide we have on the screen?
MR SHIEH: Yes, this is the measurement, because if you look at the bottom right-hand corner, you see "Total sample number for analysis", one says "11", the other says "7".
COMMISSIONER HANSFORD: Okay.
MR SHIEH: We are talking about the same 11 spots, but since every spot has two sides, on the capping beam side, as Prof Yin accepts, all 11 are valid readings, but on the slab side, four readings are invalid, so there's only seven.
COMMISSIONER HANSFORD: Yes.
MR SHIEH: But the defects are two.
Can I then invite you to look at the details. Look at the capping beam side, the green column, "Number of exposed threads", 10 to 11 -- look at the first item, 10 to 11 exposed threads, it's not defective. When it gets to 17 to 18, it becomes defective. Then further down, in specimen number 6, it's 15 to 16. Then it's

Page 115

defective.
A. Mmm.
Q. So can you help me. What kind of criteria is adopted for passing and failing in terms of exposed threads?
A. Again, you are asking an engineering problem; right? This defective/non-defective is not my expertise, even not for this capping beam; for all data, it's not my expertise.
Q. I'm just curious that all along we have been hearing two exposed threads or more than two exposed threads, then you have failure, but here we have 10 to 11 and yet --
A. I can give you my understanding. We have two types of rebar.
Q. Yes.
A. One type of rebar is 40 millimetres, the other type of rebar is 80 millimetres.
Q. Yes.
A. So the 80 millimetres actually can have a tolerance of two threads that can go to 88, and 40 can have one tolerance can go to 44. That's my understanding; there are two different types of rebars. That's why you can see there are so many threads out.
Q. If there are many threads exposed, you are saying that it could be because it's a longer thread to begin with, it may be a type B to begin with?

Page 116

A. It has to be; right? It has to be, otherwise there is no chance you can have 40 millimetres there. But anyway, it's just a simple calculation. I have no expertise on all this definition of "defective" or "non-defective". I was given the number 2 out of 7, 2 out of 11; I did my calculation shown in my report. This table I have never seen before.
Q. Okay. But from your educated guess, the reason why -- the high number of exposed threads which are not defective could be because the rebar was longer to begin with; correct?
A. Can you repeat your question again? You said the larger number of exposed threads --
Q. 10 to 11 exposed threads --
A. Yes.
Q. -- counted as not defective?
A. Yes.
Q. Whereas earlier on, when we discussed the acceptance criteria, you remember --
A. Yes.
Q. -- no more than two and must be 40 millimetres embedded.
A. Yes.
Q. In that situation, we apply the two exposed threads scenario; correct?
A. I honestly --

Page 117

Q. So that's why I asked you why suddenly here we jump to 10 to 11?
A. As I said, there are two types of bars. As far as I understand, this is possibly the longer threaded bar, 88 millimetres. But again this is not my expertise. I don't think I'm in a position to answer why there's a defect, not defect.
Q. But following our logic, if you factor in the existence or the possible existence of type B bars --
A. Yes.
Q. -- so you allow for more exposed threads, the question then arises why does that logic not feed into the acceptance criteria for the other bars?
A. For type A or --
MR KHAW: If I may just interrupt, if Mr Shieh is comparing, for example, the 10 to 11, "Number of exposed threads", under the column of "Capping beam side", with for example 44.5 millimetres, that comparison may not be meaningful, because, one, it's under the column of "Capping beam side". Another is the "EWL slab side". So you are not comparing apples to apples.
MR SHIEH: Well, I can only work on this document, because the other curiosity about this document is, Prof Yin -- help me if you can, but if you can't, just tell us -- for the capping beam side, the test criteria, the

Page 118

passing criteria, seemed to be the number of exposed threads; right? Because if you look at the green column, "Capping beam side", "Number of exposed threads", "10-11", "10-11", and then you look at the other green column on the right-hand side, "Status for statistic analysis": "Not defective", "Not defective"?
A. I see that.
Q. So it seems that for the capping beam side rebars, the pass/fail criteria utilises number of exposed threads; yes?
A. Yes.
Q. Whereas if you look at the slab side, the criteria seems to be enhanced PAUT engagement length.
A. Yes.
Q. Do you understand why that is so?
A. I don't know. I'm not an engineer.
Q. Yes.
A. I'm sure some professional engineer did the test and they came up with the number. I was given the number.
Q. Okay. Let me just ask you a question standing back. In the calculation of the strength reduction factor, let's look at your formula that you took us to just now. The number two, number of defects 2 --
A. Yes.
Q. -- over 7 featured quite a lot; right?

Page 119

A. What do you mean "feature"?
Q. 2 over 7, it featured as part of the equation that you used to derive the strength reduction factor; correct?
A. What's the meaning of "feature"?
Q. It appeared in the equation.
A. Okay, yes, of course.
Q. 2 over 7; correct?
A. Yes.
Q. Now, you would accept that seven is a relatively low number of specimens?
A. When you say "relative", relative to what?
Q. The total number of couplers in the entire EWL slab.
A. The total number of EWL slab is 90; right?
Q. Mm-hmm.
A. It's 90, and now you have seven, and on the other side you have 2 out of 11.
Q. Yes. The question I have is this. You accept that this problem about couplers appearing in a panel with capping beam details is something that the workmen stumbled across as and when they did the opening-up; correct?
A. That's what I was told.
Q. That's what you were told.
A. Yes.
Q. So, at the planning stage, you know, when you were theoretically planning all this --

Page 120

A. Yes.
Q. -- the model had not taken into account the need to separate between, "For couplers in panels with capping beam details, let's do it this way, let's do the sampling one way; for those in panels without capping beam details, let's do the sampling some other way" -- this was actually not taken into account at the original planning stage; correct?
A. Yes.
Q. My question is this. Is it not possible for the purpose of working out the strength reduction factor not to zoom in and highlight the strength reduction factor attributable to those panels with capping beam details? Rather, you look at the entirety of the EWL panel couplers as a whole? That's the question.
A. I don't get what you are asking.
Q. If you zoom in on the capping beam locations --
A. Yes.
Q. -- and you look at 2 defectives over 7 --
A. Yes.
Q. -- that faction would stick out and be factored into the equation that you used to derive the 60-odd per cent?
A. Yes.
Q. Whereas if you don't just single out the panels or the couplers, on the locations with capping beam detail, as

Page 121

a stand-alone population, you simply group those couplers as part of the couplers in the entirety of the EWL slab, you could very well get a completely different result; do you accept that?
A. Please help me to understand your question.
Q. Mm-hmm.
A. So what we did here is we take capping beam alone and try to work out the capping beam -- the couplers involved in the capping beam side, what is the defect rate, and originally we had this calculation for EWL and NSL, because the other side of the coupler was embedded in the D-wall. You don't open the other side of the coupler. You only open one side of the coupler and look at whether it's properly installed. But it happens that for the capping beam, both sides were exposed. Then I was approached immediately, what kind of statistical method could be used. It basically opened another can of worms and we derived the probability formulas and we said, okay, the capping beam, we focus on capping beam alone, and that's the formula we used and we derived this 68.3 per cent.
Q. Dr Wells' complaint, which you have not spoken to, is in his paragraph 4.40, of Dr Wells' report.
A. Okay. Yes.
Q. 4.40(b) --

Page 122

A. Okay.
Q. -- he said:
"It is not clear ... why data relating to the EWL slab side does not also use the main EWL data set, as doing so would greatly increase the confidence in the results as well as overcoming some of the mistakes made (by assuming a large sample approximation, when the sample size was actually very small) ..."
A. I understand what he is talking about.
Q. You understand. That's the point I'm making.
A. Of course I understand. Then what is your question?
Q. What do you say about his complaint, about not taking into account the EWL data set as well? In other words, why zoom in on the number of couplers in locations with capping beam detail and to single out defective couplers in that location as some kind of a separate population, to do your sum about 2 over 7?
A. Yes. Let me explain.
Q. Do you understand?
A. I understand perfectly, yes. I understand what he was trying to say. So, basically, on one side, which is capping beam side -- okay, capping beam side -- the other side is the slab side. I think what he was trying to say, on the slab side, you have more samples, but on the capping beam side you don't. Capping beam side --

Page 123

I forgot, seven is on the capping beam side or 11? 11 on the capping beam side? Okay. Thank you.
So capping beam side, you have 11 samples. That's it. You cannot enlarge that sample size any more. But he had some valid argument, "Okay, on the slab side you use seven only. Why don't you use all the slab side couplers?" Actually, what I did, I did as what he said, and the results reduced a certain extent, but not much. I did do what he said. I don't have the results here. But -- because, you see, Dr Wells did a whole lot of calculations, and I could verify some of the calculations by his suggestion, and actually I did this kind of sensitivity analysis, what I call.
But using two out of 11 and two out of seven, because we consider the capping beam section is special, somehow different from the other side of the EWL, and whether this is a valid approach or not, I think you have to put an engineering consideration into this problem. On the engineering side, the engineers, they say, "You should treat them separately."
COMMISSIONER HANSFORD: Is that what they did say? Did they say you should take them separately?
A. Based on my calculation, clearly we already agreed, otherwise I wouldn't do this two out of seven, two out of 11. Clearly, we had agreement that this would be

Page 124

treated as a separate population, because this capping beam is -- you had two sides of the coupler, not like D-wall, you only have one side exposed.
Now, certainly you can have argument, pool them together. But even if you pool them together, on capping beam side, all the samples you have is 11. That's the only sample you have.
COMMISSIONER HANSFORD: I'm not sure what's so special about the capping beam side.
A. I don't either. I don't know what's so special about the capping beam.
COMMISSIONER HANSFORD: To my mind, it's about screwing in bars into couplers, irrespective of where they are.
MR SHIEH: Except that in a normal case, you screw one side; in a capping beam situation, you screw the other side, so I don't know.
A. Yes, I totally agreed with you, I tried to look up capping beam on Google, tried to view some YouTube. It doesn't help me.
CHAIRMAN: That's a pretty dangerous thing!
A. It doesn't help me, so I have to rely on the engineer's input.
MR SHIEH: Thank you very much. But at the planning stage nobody thought somehow this should be a special population, some special formula should be derived to

Page 125

cater for two separate populations; it wasn't there at the planning stage, correct?
A. Yes.
Q. Just now, you said, in addressing a possible layman reaction, why 60-odd per cent, so big, and then you try to do some kind of a cross-check by utilising the 30-odd per cent strength reduction factor for EWL and NSL?
A. Yes.
Q. You used a factor of 1 minus 30-odd per cent, and then multiplied by the other bracket, 1 minus 30-odd per cent, and then 1 minus, I think?
A. Yes, exactly. Good memory.
Q. I haven't got the slide here, but to save time I'm not going to call that up. You remember which slide we're talking about?
A. Yes, I remember.
Q. But that cross-check is only valuable if the underlying 30-odd per cent holds good. In other words, if 30-odd per cent --
A. 30 what?
Q. "30-odd".
A. "30-odd"?
Q. 30-odd per cent strength reduction for EWL and NSL.
COMMISSIONER HANSFORD: "30-odd" means approximately 30.

Page 126

MR SHIEH: Yes, 30-something per cent.
A. You should tell me exactly the number.
MR SHIEH: It doesn't matter. Because for both NSL and EWL, it's 30-something per cent.
A. Yes, 30-something.
Q. But if for every 30-something per cent it gets knocked down to, let's say, 10 per cent, then the resulting number would change. That's a matter of simple arithmetic.
A. Of course.
Q. So the value of this cross-check is only as good as the underlying percentage of the strength reduction factor applicable to the EWL slab and the NSL slab; correct?
A. That part of the calculation is trying to explain to laymen why 68 per cent seems to be a high number -- how, why you can come up with 68 -- so I mentioned clearly it's not rigorous, it's loosely speaking, gives you intuitive reason, and as I said, EWL, we estimated -- yes, it's right here.
Q. Yes, 1 minus 0.366.
A. Yes. You see EWL defect rate is 36.6 per cent.
Q. Yes.
A. Let's just pretend to take this 36.6 per cent, apply this number to both left and right.
Q. Yes.

Page 127

A. Pretend. Even though you pretend to do this, strictly speaking, this is not correct.
Q. But you are saying it's loose, not rigorous, just a cross-check?
A. Not even a cross-check. This just tries to help non-statisticians understand why there's 68, such a large number. This is not even cross-checking because this calculation is wrong from the beginning. I'm talking about wrong. Do you know why? Because 36.6 is not the defect rate. It's the upper bound of the defect rate of the 95 per cent confidence interval.
If you recall, in Wells' report, basically you have a stated mean defect rate plus 1.645 times standard error. There's a non-linear transformation there based on the normal curve. That's where 1.645 comes from. You have this kind of non-linear transformation.
So that's why I say this whole calculation is just for laymen to have a sense why it can go up as 68.3 per cent. This is not even cross-check.
Q. Thank you.
Can I move on to my final topic. There's one point in the verification report about the untested rebars --
A. Okay.
Q. -- that I wish to talk about with you.
A. Yes.

Page 128

Q. Can you look at your report, at 6.3.2. That's for the verification, that's in COI 2.
At paragraph 6.3.2 ...
A. Yes.
Q. In fact, this section starts at 6.3.1:
"As explained in section 4.3 of the verification report, in order to determine the effect of the 7 per cent of untested rebars on the completed structures, the testing records of MTRCL's HOKLAS laboratory were used as a reference. Over the past 9 years ..."
That was 2010 to 2019; right?
A. Yes.
Q. "... about 110,000 rebar samples were tested at MTRCL's laboratory and out of which 55 samples failed the test. These 55 samples are divided into two groups, ie samples with a bar diameter equal to or greater than 16 millimetres and samples with a bar diameter of less than 16 millimetres. For the former group, the worst case failure gives a tensile strength reduction of 4 per cent, ie the measured tensile strength of the worst case is 4 per cent less than the design tensile strength. For the latter group, the worst case failure gives a tensile strength of about 13 per cent lower than the required design strength. In other words, these

Page 129

4 per cent and 13 per cent strength reductions represent the worst case scenario of the 55 failed samples. They are the extreme failure cases by tests and were not derived from any statistical analysis. I note that these two strength reduction factors were used in the structural review, by assuming that the said reduction factors apply to all rebars used in NAT, SAT and HHS, in order to assess if the completed structures could accommodate such strength reduction.

This is a sensitivity analysis by plugging in the worst case scenario based on the information from past experience rather than a statistical analysis."

Do you see that?

A. Yes, I see it.

Q. So, basically, what's happening is one wants to find out the significance of the 7 per cent untested rebar.

A. Yes.

Q. So one says, for the past nine years, the entirety of all bars tested by MTR, for whatever project, whatever manufacturers, let's lump them together. There were 55 failures so let's assume that the untested rebars are going to be -- the fate of the 7 per cent untested rebars would more or less follow the worst-case scenario demonstrated by this result, 55 out of 110. It's based on an assumption; correct?

Page 130

A. Based on what assumption?

COMMISSIONER HANSFORD: 55 out of 110,000, Mr Shieh.

MR SHIEH: Yes, 55 out of 110,000.

COMMISSIONER HANSFORD: The transcript says 55 out of 110.

MR SHIEH: 110,000.

So it was assuming that the untested 7 per cent of the rebar has attributes or qualities which are comparable to the rebars that have been tested by the MTR for the past nine years, forming that 110,000 population for sample?

A. Yes. My understanding, the 7 per cent untested rebars, that basically means untested after the 7 per cent rebars being delivered to the site. So my understanding is there is CS2:2012 clearly states there's manufacturer's test, there is the purchaser's test. That's basically on-site delivery test. And this 7 per cent is after the batch delivered on site but they didn't test this 7 per cent. And based on this laboratory test, there are 110,000 rebars tested in the past nine years and 55 samples are defective.

Q. Right.

A. Yes, that's all the information.

Q. And, of those defective samples, they are of different diameters?

A. Yes, that's --

Page 131

Q. So two categories; right? One would be greater than or equal to 16, and the other would be less than 16; right? Two families; right?

A. Yes. Yes.

Q. So, within each family of failures, you pick the worst case; is that what has been done? You pick the worst example for each family; correct?

A. I think this is in the verification report.

Q. Yes. Let's look at that. 4.3 of the verification report.

A. Yes, this is in the verification report.

Q. Yes. Let's look at that. BB16, page 9977. 4.3.2. It's actually similar to what you have said in your report.

A. Yes.

So this is exactly -- I'm copying what they were saying.

Q. Yes.

A. I claim this is not a statistical issue.

Q. You assume that the 7 per cent untested rebars for this project would be -- you are assuming that the untested 7 per cent has a quality equal to the worst case situation of the failed cases within the past nine years. Is that what you're saying?

A. I think what I was -- just trying to repeat what is

Page 132

written in the verification report.

Q. Yes.

A. And I emphasise this is not a statistical problem, there is no statistical model.

Q. I know. There are 55 failed samples out of 110,000 samples tested by MTRC's laboratory; yes?

A. Yes.

Q. Those 110,000 samples could come from any project, of any manufacturer; right?

A. I don't know.

Q. You don't know?

A. Yes.

Q. Fine. Basically, is it the purpose of the exercise, first of all, to say within these 55 failed cases, let's say this many belong to this diameter and that many belong to the other diameter, so you divide them into two families, correct, according to diameter; yes?

A. Clearly that's what is written here, yes. 16 millimetres is the threshold.

Q. Yes. And from each family, basically, the exercise -- tell me, because you said so in your report and I'm asking whether this was in fact what was done -- that is, for diameter of 16 or above, you go back to look at the failed cases of 16 or above, and you look at the worst case out of those failed cases and check how many

Page 133

per cent strength reduction that adds up. Is that the exercise performed?
A. No. I didn't do this exercise. I simply repeat what is written in the verification report, and I said this is not a statistical issue. That's all I said.
Q. Okay. So somebody decided to assume that the 7 per cent untested rebars would be of a quality equivalent to the worst example of the failed cases tested by the MTR in the past nine years?
A. Yes, but I just -- as a layman, about all this rebar testing or whatever being written in this paragraph, I'm just wondering why those 7 per cent never been tested.
It's clearly saying manufacturer's test, purchaser's test, why those 7 per cent not tested, and now you want to do back-calculating. Again, all this back-calculating performed by Dr Wells, it's not so meaningful. He was trying to back-calculate whether the random sample is genuinely random. I can tell you any sample can be a valid random sample. If I toss a die, you observe ten 6s -- ten times 6 numbers, it's a rare event but it may occur. You do Mark 6, any number can come out. It's a valid random sample.
The only difference, probably the chance is slim to observe those rare events, but you can't question whether it's random or not random.

Page 134

So I think all this back-testing or back-calculating those probabilities, it's not meaningful. It's just my understanding.
Q. You are not here to ask me questions.
A. Okay.
Q. But I can answer you by telling you that the reason why things are not tested is a separate exercise before this Commission, and the Commissioners will consider the answer to that question, but that is not a statistical matter so let's leave that to one side, why those were not tested.
A. Okay.
Q. My final question to you is -- you said that the assumption of the worst-case scenario is not a statistical matter; correct?
A. It's not -- I clearly said it's not a statistical issue.
Q. It is not something you recommended to be done; correct? So you played no part in adopting that approach in the verification report; correct?
A. If it's not a statistical issue, I have no say about it, the paragraph there.
MR SHIEH: Thank you very much. I have no further questions.
WITNESS: Thank you very much.
MR BOULDING: No questions from MTR, sir.

Page 135

MR PENNICOTT: Sir, I have a few questions. It probably won't take more than five minutes or so, so if we can ask the transcript writers to bear with us, perhaps we can just plough on, unless you really --
CHAIRMAN: Yes, that sounds best.
Would that be all right?
COURT REPORTER: Yes.
MR PENNICOTT: Thank you.
WITNESS: Excuse me, because I think I want to go to the restroom.
CHAIRMAN: Sorry.
WITNESS: Is that okay? I will come back right away.
CHAIRMAN: Ten minutes.
WITNESS: Thank you very much, Chairman.
(4.08 pm)
(A short adjournment)
(4.20 pm)
Examination by MR PENNICOTT
MR PENNICOTT: Thank you, sir.
Good afternoon, Prof Yin.
A. Good afternoon.
Q. My name is Ian Pennicott. I'm one of the counsel to the Commission. Thank you very much for coming along to give evidence to the Commission. I don't think that's been said to you by anybody yet.

Page 136

Prof Yin, I just have a few questions for you.
A. Okay.
Q. And they really all relate to one particular topic.
A. Okay.
Q. Which is the combined defective rate in relation to the capping beam coupler connections, a matter that Mr Shieh asked you some questions about earlier and I just want to get a bit more clarification, if I may.
A. Okay. Yes.
Q. Prof Yin, the first point really is a pure factual point which I would ask you to clarify, if you would, please. Could I start by asking you to look at a document you may not have seen before, but it's the MTR report on statistical analysis that they produced for the purposes of the Commission. It's in ER1, I think at tab 11.
Do you see that, Prof Yin? Is this a document you've seen before?
A. I don't remember.
Q. Okay. Not to worry. Let's just plough on.
If you could go to paragraph 15, please, internal page 7.
A. Okay.
Q. As I say, this is a document -- so you are not misled by anybody, particularly me -- prepared by MTR.
A. Okay.

Page 137

Q. What they say is, at paragraph 15:
"Broadly speaking, the statistical analyses adopted in stage 2b of the holistic report include:
(1) Binomial statistical analysis; and
(2) The Formula" -- capital F, a point mentioned earlier by Mr Shieh -- "(as defined at paragraph 43 below)
to cater for the different situations as explained below."
Do you see that, Prof Yin?
A. Yes.
Q. Can we then go to paragraph 43 -- perhaps paragraph 42, just to get the introduction to this. The report says:
"In mid-June 2019" -- I'm reading from the bottom of page 17, Prof Yin, paragraph 42.
A. Okay.
Q. "In mid-June 2019, MTR proposed using binomial analysis to calculate the defective rate for each of the EWL slab side and the capping beam side coupler engagements, followed by a probability analysis to calculate the combined reduction factor. The task force group commented that MTRC's proposed analysis was not acceptable from a statistical perspective."
Then the more important part, 43:
"Eventually, a formula as shown in appendix II (the

Page 138

'Formula') was shared with MTR, which MTR understood came from the government's statistical advisers and would account for: (i) the combined defective rates of the coupler connections at both the slab side and the capping beam side; and, (ii) the small sampling size at the capping beam area."
Do you see that?
A. Yes.
Q. If you go to what's described as appendix II -- it's in the same file, towards the end, right at the end -- you will see there, on a document that is probably now familiar to you because it was shown by Mr Shieh to you earlier, albeit in a different place; do you see that, Prof Yin?
A. Yes.
Q. So it's the one that has the combining calculation, if I can call it that --
A. Okay.
Q. -- but a document produced, it is said, at this stage by MTR.
A. Yes.
Q. Could you then please be shown the witness statement of one of the MTRC's witnesses, that's Mr Ng, which I think is in B21/26698.
A. Okay.

Page 139

Q. If we could go, please, to page 26703, paragraph 18. What Mr Ng says in paragraph 18 -- and Mr Ng gave evidence earlier this week, Prof Yin -- he says this:
"As set out in paragraph 15 of the COI ... stat report" -- that's the one in paragraph 15 I took you to just a moment ago -- "broadly speaking the statistical analyses adopted in the holistic report include: (i) binomial statistical analysis; and, (ii) Prof Yin's suggested Formula (the 'Formula')."
Do you see?
A. Yes.
Q. Perhaps I may be forgiven for thinking that Mr Ng was telling us that appendix II that we looked at a moment ago, in the report, was in fact your calculation, but as I understood your answers to Mr Shieh, you have not -- you certainly weren't, you say, responsible for preparing that document, indeed I think you said you had never seen it before; is that right?
A. I never seen this document before, but the calculation as he said I suggested -- but, you see, I didn't prepare anything here. Maybe some suggested how to calculate that I explained and he worked out the detail. I don't know. I wouldn't use this kind of symbol in my derivation. You look at my derivation, it's very mathematical. This is very like English writing, like

Page 140

essay writing. I wouldn't use three letters to denote one symbol. So this whole thing here, the appendix, I probably explained to him how you are supposed to do the calculation, but I didn't write the whole thing.
Q. All right. We've seen, Mr Shieh took you to it, that in your report for the Commission you have set out your calculation, and we've looked at that.
A. In my report?
Q. Yes, in your report.
A. Yes, I had my own calculation.
Q. That's right.
A. And I verified everything in my report, calculated.
Q. Of course you've got your calculation in your report and that makes no reference to the 237 and the split between 175 and 62, there's no reference to Qa and Qb?
A. Yes.
Q. Did you ever -- you've obviously clearly had a discussion with MTR --
A. Clearly, yes.
Q. -- about this formula. Did you ever supply them with an actual formula, or did you just discuss the parameters of it and how to do it?
A. I had a discussion with MTR. I explained the problem. They encountered two sides of one coupler. They were confused. They approached me and my team and we spent,

Page 141

I don't know, the whole afternoon to discuss the whole thing. I explained it on the board, what you are supposed to do, and then that's what he did.
Q. Right. Okay. But your evidence to the Commission -- and you've said this a couple of times now so I'm sorry for repeating it -- but you never carried out a calculation that utilised the 237, the 175, 62, and the Qa and Qb; that's not your approach?
A. That's not my approach. Let me clarify again: using Qa and Qb in this derivation is for different purpose. I sort of understand what he was trying to do. He was trying to compute an overall defect rate, overall, and what I did, in my report, is focusing the capping beam only.
Q. Right.
A. And I had a computer code, programmed everything, verified all the results. You can check whether they did the same thing too.
Q. When you were asked to look at this combined calculation, were you ever given any indication as to what the purpose of that calculation was?
A. Yes. My understanding is, as I said, a coupler would function if both ends butt-to-butt, and now you have four possibilities: both sides pass, one side passes, one side fails, or vice versa, or both sides fail.

Page 142

That's my understanding. And to calculate the failure rate for coupler with both sides being considered, you have to go through this kind of probability derivation.
Q. Were you ever told how this calculation might be used in terms of extrapolating the results to area A?
A. No. I was not aware of this at all.
Q. Right. To be fair to you, Prof Yin, in your reports to the Commission, you make no reference to the extrapolation exercise to area A. However, in your slides this morning --
A. Yes, I remember.
Q. -- you do make reference to it.
A. Actually, what I put there is -- because this has been going on, discussed so many times, about area A no coupler, why, whether you can being extrapolate -- I've been thinking all these days should we extrapolate or whether you call it an extrapolation? It's more an engineering problem. They can decide if they want to extrapolate it or not. But if you ask me as a statistician whether I would agree with extrapolation, I would say so. Even though it's more engineering problem, you can discuss with engineers whether -- because they know the structure more than I do. There are two things I really have no idea. I only been given the information they are similar configurations, same

Page 143

contractor, similar workmanship, that's all the information I have. How close they are, I don't know.
Q. I understand your position, Prof Yin, and I just want to press you a little bit further on it --
A. Okay.
Q. -- just to see how genuine this is an engineering manner rather a statistical matter, and I think you have fairly said, at least on a couple of occasions now, that as you view it, it's primarily an engineering matter; would you agree with that?
A. Yes, I agree with that.
Q. Can I ask you to be shown your slide 20.
A. Okay.
Yes.
Q. Perhaps the first thing to note is the heading, Prof Yin.
A. Yes.
Q. You say "Possible" --
A. And question mark.
Q. -- and you have a question mark.
A. Yes.
Q. That's right. As I think you have just indicated, this is a question that has been perhaps running through your mind --
A. Yes, I've been thinking about it.

Page 144

Q. That's fine. So far as the words in red are concerned, Mr Shieh I think asked you about one or two of these matters, but when you say, for example, "similar configurations", you mean similar configurations of the rebar and the coupler connections; is that what you mean?
A. "Similar configurations", yes, you can think that way.
Q. "Same contractor" -- perhaps you mean the same sub-contractor, the sub-contractor responsible for installing the rebar?
A. Yes. I was given this kind of information. I don't know. I couldn't verify this.
Q. Right. You say "similar workmanship".
A. Because of same sub-contractor.
Q. Okay. What other factors might be important, do you think? The same actual workers doing the work?
A. Let me put it this way. If you, suppose, need a kidney transplant, you need to find someone who has a lot of similarities to you in order for this transplant to be working. I'm using medical example again. So you've got to find someone who matches your body so that your immune system would not kick that thing out, because there is a danger thing that thing does not work in your body and the patient would die very quickly.
Q. Yes.

Page 145

A. So that's called a GvHD, graft versus host disease, people will die in a couple of weeks if that thing is not accepted by the body.

So similar, basically, you -- how to say -- you find some people who are the same blood type, similar genetics, all these things similar, and then you can do a successful donor transplant.

Q. In a nutshell, Prof Yin -- perhaps it's just a matter of common sense -- the more similarities you can find the more likely it is that the extrapolation is justified?

A. Yes.

Q. All right. Now, you've referred to similar configurations, same sub-contractor, similar workmanship. What about if I told you that the work, the relevant work, that is the connections in area A and area HKC were carried out a year apart, would that be relevant, or is that not a statistical matter?

A. I think it's not a statistical matter. It's just a common-sense matter. I think anybody can have their own view. This is not a -- statistical matter, you have to give me the numbers and I do the calculation for you. If you give me a sentence everybody can interpret it differently. I need the actual number and then I can program it and do calculation for you. But if you just ask me a very general, broad question, I don't think it

Page 146

is a statistical question.

Q. We were looking for similarities and one similarity might be they were carried out at roughly the same time. That might be another factor. And I'm just indicating to you that as a matter of fact, the area A work was carried out in May and June 2015, and the HKC work was carried out in July and August 2016, over a year later. I just wondered whether you had any view as to whether that might be relevant.

A. I cannot answer this question. I don't know.

Q. All right. So ultimately, Prof Yin -- my last question, really -- you are, as I understand it, content -- you have asked yourself the question, you have put a few thoughts down --

A. Yes.

Q. -- and ultimately you are prepared to leave this matter to the engineers?

A. Yes.

MR PENNICOTT: Thank you very much. I have no further questions.

Re-examination by MR KHAW

MR KHAW: Prof Yin, it's been a long day for you. I only have two questions --

A. Okay.

Q. -- which arise from Mr Shieh's discussion with you

Page 147

today.

Perhaps I will first provide you with the background regarding your discussion with Mr Shieh and then I will ask you to clarify a few points.

A. Okay.

Q. Now, you recall that before our lunch break today you were asked by Mr Shieh regarding who decided the binomial analysis; do you remember that?

A. Yes.

Q. Then you were also asked about what information was actually placed before there was a decision on a binomial analysis; do you remember that?

A. Mm-hmm.

Q. Now we all know that in terms of the results regarding the tests for coupler connections, we have this classification of only two types of results: pass and fail.

A. Yes.

Q. I just want you to clarify this. If you can take a look at your own report, first report.

A. Okay.

Q. If I can ask you to take a look at page 17, paragraph 3.2.2.

A. Okay.

Q. Here you said:

Page 148

"In the design stage of the holistic proposal, I verified the suggestion using a binomial analysis by MTRCL."

Pausing here, I think we have seen evidence elsewhere that the binomial analysis was proposed first by MTR; right?

A. Yes.

Q. So that you have no dispute about; right?

A. Yes.

Q. But you said you verified the suggestion using a binomial analysis by MTRCL. So, at the time when you were doing your verification for the suggestion of this binomial analysis, were you actually given this information, namely the information that the results will be classified into two types, pass and fail only? Were you given to understand that at that time?

A. You mean based on the criteria, based on the two criteria? I was not given any criteria.

Q. Right. Let's not talk about the details about the criteria first.

A. All right.

Q. 37 or 40mm, et cetera. Let's not talk about that.

A. Okay.

Q. Just the classification of pass and fail. Leaving aside the details of the criteria.

Page 149

A. Okay.
Q. That piece of information, ie pass and fail, the end result will only be that classification. That piece of information, was it given to you at the time when you verified whether binomial analysis should be used or not, or afterwards?
A. I don't quite get the question, because binomial is about "yes" or "no", pass or fail. That's what binomial means.
Q. Yes.
A. Just like you ask me a question "yes" or "no", I have to say "yes" or "no". That's binomial. So we've been using binomial for the whole day in this courtroom. So binomial is a very standard approach, and for me to look at the problem, the first natural approach for me is binomial approach, and I have already stated so many advantages about why we use it. If we try to use multinomial, continuous outcome, things could become complicated, even become infeasible.
Q. Right. So, at the time when you were considering the suggestion of a binomial approach, were you actually given the details regarding the acceptance or rejection criteria?
A. No.
Q. Thank you.

Page 150

The next question is about one part of Dr Wells' report. If you could take a look at his report, his report for the Original Inquiry, page 11, paragraph 4.40.
A. Okay.
Q. I think Mr Shieh took you to subparagraph (b); do you remember?
A. Yes.
Q. It says:
"It is not clear to me why data relating to the EWL slab side does not also use the main EWL data set, as doing so would greatly increase the confidence in the results as well as overcoming some of the mistakes made (by assuming a large sample approximation, when the sample size was actually very small) ..."
Do you remember that?
A. I remember, yes.
Q. I think, in answer to Mr Shieh's question in this regard, you told us that you had in fact carried out verification or carried out calculations --
A. Yes.
Q. -- on the basis of the entire EWL slab date for the EWL slab side; right?
A. Yes.
Q. Then you told us that the difference was minimal?

Page 151

A. It's not dramatic.
Q. Can you recall the --
A. I cannot recall the exact number, but I can tell you this. First, Dr Wells' argument is reasonable. He says the sample size is small and I use normal approximation. His argument is reasonable so I went back and used the bootstrap. As I said, I used the bootstrap and verified the answer would be very close, so it's kind of reassuring my own calculation.
Secondly, I also did some sensitivity analysis by using all the EWL data, on the EWL slab side, and the other side use capping beam side, still two out of 11. I did this calculation, I don't recall exactly the number, but the number is not dramatically decreased.
Actually, he, Dr Wells, is correct in the sense that if you have a larger sample size, certainly you have higher accuracy. That's very much common sense in statistical perspective.
Q. Now, in terms of the verification that you have done, the calculation you have done, by taking into account the entire EWL slab data, would you be able to tell, in terms of some degree of magnitude, the percentage? Is it possible?
A. It's possible. If you allow me to use my -- look at my laptop, my laptop has the result. Am I allowed to check

Page 152

that result?
Q. If you can, please.
CHAIRMAN: Certainly.
A. Okay. Let me open my laptop and I can tell you right away.
How do you want me to show the result to everyone? Can this be shown to other people? Because I've found the table I produced.
CHAIRMAN: You can just tell us what the result is for the moment, and those who want to have a look --
MR PENNICOTT: Show it to Mr Khaw first. (Handed).
A. But I think he may have a hard time to understand all my notations.
MR KHAW: I'm sure.
A. It's just a table for myself --
MR PENNICOTT: Presumably, you have been teaching him a bit!
A. Do you need me to help you understand?
CHAIRMAN: Mr Khaw, would you like the professor to assist you for a moment?
MR KHAW: It would be helpful, I'm sure.
A. I can explain the numbers a little bit to you and then you can --
COURT REPORTER: You need to speak into a microphone, any microphone.
CHAIRMAN: Any microphone.

Page 153

A. Okay. So look at the top row. It's "2", "7", "2", "11". That's our original calculation; right? So this second row is "25", "90", "2", "11". The number dropped from 68.3 per cent to 56.1 per cent. If I use 90 couplers as the total sample size, 25 failures on the EWL slab side.

If I further use 183, 48 failures, it will keep dropping to 54.4 per cent. So, basically, the defective rate dropped from 68.3 per cent to 56.1 per cent.

For the 12 points --

COMMISSIONER HANSFORD: Sorry, was it 54.4 or 56.1?

A. 56.1.

COMMISSIONER HANSFORD: What was the 54.4?

A. That's if I combine all the samples, including I think NSL as well.

This is EWL. If I combine EWL and NSL, all samples together, they will be further drop, but the drop is to 54.4.

So this is basically trying to verify Dr Wells' argument, "Why don't you combine all these samples instead of just using 11?"

MR KHAW: So 56.12, that is the result we get after taking into account the EWL slab entire data; right?

A. 56.1, yes, EWL entire data. That's basically 90 samples, 25 failures.

Page 154

COMMISSIONER HANSFORD: And 54.4 if you then add the NSL data; is that correct?

A. Yes.

COMMISSIONER HANSFORD: Would it be helpful -- please tell me if it wouldn't -- if we had a print-out of that page?

MR PENNICOTT: We can certainly ask for a print-out of that page, yes.

COMMISSIONER HANSFORD: Thank you.

MR KHAW: I was about to suggest that, yes. I think we will make the arrangements.

MR PENNICOTT: Not necessarily now.

COMMISSIONER HANSFORD: Not immediately, no.

MR KHAW: I have no further questions.

CHAIRMAN: I'm just thinking of the print-out. It could be done by the professor and he can arrange to deliver.

MR PENNICOTT: I'm sure the government can arrange for it to be done, sir.

CHAIRMAN: Yes, or maybe it can be done this afternoon here. I don't know if you have anybody capable of doing it.

MR KHAW: Yes.

CHAIRMAN: Good. Thank you.

Mr Pennicott, anything further --

MR PENNICOTT: Not from me, sir.

CHAIRMAN: -- this afternoon? Good.

I think then that we are adjourned through until

Page 155

next Friday.

MR PENNICOTT: That's correct, sir. Yes, when we will commence the hearing of the further project management experts' evidence.

CHAIRMAN: Yes. Good. Thank you.

Further to what happened this morning, the meeting --

MR PENNICOTT: Yes, sir.

CHAIRMAN: -- I unfortunately won't be here for a little while. I have a conference in Singapore to go to on Monday, but if you could perhaps, Mr Pennicott, when you have time, if you have time, perhaps speak directly to Mr Boulding, Mr Khaw and others, just to see what would be an advisable way forward from your perspective.

MR PENNICOTT: Yes, sir.

CHAIRMAN: And bearing in mind Leightons of course.

MR PENNICOTT: Of course.

CHAIRMAN: It's important that no party is prejudiced.

MR PENNICOTT: Yes. We will try and put something very briefly in writing over the weekend or early next week.

CHAIRMAN: Thank you.

MR PENNICOTT: Which we may circulate to everybody, and then we will decide the best way forward in terms of formalising the position.

CHAIRMAN: Good.

Page 156

MR PENNICOTT: Sir, could I just mention, in relation to the project management experts, I think I've been keeping up with the emails that have been going to and fro today, but my understanding is that Mr Huyghe, the MTR's project management expert, is producing a further report which I think he's just been given another day's extension to produce, which I think will be therefore tomorrow evening.

CHAIRMAN: Yes.

MR PENNICOTT: And we are expecting a joint statement from the experts I think now, again with a further extension being granted, on either Wednesday but it might Thursday -- it's either the 2nd or the 3rd, I've forgotten -- but the middle of next week.

CHAIRMAN: All right. Have we cleared up the difficulty with Mr Wong, I think it is?

MR PENNICOTT: No, we don't need to concern ourselves with that at the moment.

CHAIRMAN: Good. Thank you very much indeed.

Professor, thank you very much. I know we held you this morning, but you have been of really great help. Thank you very much indeed. I just hope it hasn't been too much of a strain for you today. But thank you.

WITNESS: Thank you.

CHAIRMAN: Good. Anything further?

(The witness was released)
MR SHIEH: There is an outstanding witness from Leighton who was stood over from Monday.
CHAIRMAN: Mr Cowley, yes. I thought we were going to include him with --
MR PENNICOTT: We are, and I apologise to Mr Shieh that I forgot that. I can speak to him about logistics. On the assumption of course that Mr Cowley is available next Friday, I certainly would suggest that we call him first, but obviously that's subject to Mr Shieh telling us that he's available.
CHAIRMAN: All right.
MR PENNICOTT: Perhaps I can have a word with Mr Shieh about that when we break.
CHAIRMAN: Of course. Yes. Thank you very much indeed. Anything further? Good. Thank you all very much indeed. So we are adjourned until 10 am -- 10 am?
MR PENNICOTT: Yes.
CHAIRMAN: 10 am next Friday. Thank you.
(4.55 pm)
(The hearing adjourned until 10.00 am on Friday, 4 October 2019)

INDEX

PAGE