Extended Commission of Inquiry into Construction Works at and near Hung Hom Station Extension under the Shatin to Central Link Project ("Extended Inquiry")

Expert Report
Prepared by Barrie Wells

13 September 2019

CONTENTS PAGE

1. Introduction

1.1. I have been engaged by O'Melveny & Myers on behalf of Leighton Contractors (Asia) Limited ("**Leighton**") to provide statistical expert evidence for the Extended Inquiry in relation to the Final Verification Study Report on As-constructed Conditions of the North Approach Tunnels, South Approach Tunnels & Hung Hom Stabling Sidings [BB16/9952] ("**Verification Report**").

1.2. My relevant area of expertise is in quality assurance testing and the statistical analysis of sample based testing for the purposes of assessing conformance with standards.

1.3. I hold a B.Sc. in Mathematics and Statistics from the University of Bath, an M.Sc. from Cranfield Institute of Technology and a Ph.D. in Underground Stress Analysis from University of Nottingham, England.

1.4. I have worked in statistical analysis and modelling for numerous government agencies and commercial organisations in Australia, North America and Europe, and have acted as a visiting lecturer in geostatistics at Nottingham University. Since about 1993, I have managed an independent company in North Wales, which provides services including mathematical and statistical analysis to companies worldwide. I have also provided statistical advice to standards committees in the UK and internationally.

1.5. I have provided a copy of my CV to the Commission.

2. Scope of Instructions

2.1. I have been instructed to provide my expert opinion in relation to the topics of rebar testing and coupler connections addressed in the Verification Report.

2.2. In relation to rebar testing, I have been asked to address the following questions:

(a) **Question 1**: Based on relevant quality assurance standards, what is your assessment of the quality of the rebar based on the results of the tests conducted by the manufacturers and those conducted after delivery to the site? In particular, what level of confidence in relation to the quality of such rebar can be drawn from the results?

(b) **Question 2**: What are your comments on the statistical methods and analysis referred to in Section 4.3 of the Verification Report for the purposes of assessing the strength of the rebar used in the Project?

(c) **Question 3**: What are your comments on the conclusions drawn in Section 4.3 of the Verification Report? In particular, is it reasonable to adopt a strength reduction factor of: (i) 4% for rebar with a diameter equal to or more than 16mm; and (ii) 13% for rebar with a diameter of less than 16mm?

(d) **Question 4**: Based on relevant quality assurance standards, how would you assess the adequacy of the testing conducted on the batches of rebar that were used in the Project?

2.3. In relation to coupler connections, I have been asked to comment on the statistical approach and analysis adopted in the Verification Report (including in relation to sections 3.2.12, 4.2.6, 4.5.1 and 4.5.2 of the Verification Report).

3. Rebar testing

Question 1

Based on relevant quality assurance standards, what is your assessment of the quality of the rebar based on the results of the tests conducted by the manufacturers and those conducted after delivery to the site? In particular, what level of confidence in relation to the quality of such rebar can be drawn from the results?

3.1. The quality assurance standards relevant to rebar testing are "Construction Standard CS2:1995 Carbon Steel Bars for the Reinforcement of Concrete" ("**CS2:1995**") and the updated standard "Construction Standard CS2:2012 Steel Reinforcing Bars for the Reinforcement of Concrete" ("**CS2:2012**").[1]

3.2. There are many points of detail not addressed directly in either CS2:1995 or CS2:2012 and many requirements are stated without derivation, reference or context. It is therefore necessary, if seeking either context or clarity on points of detail, to refer to the over-arching principles set out in international standards, of which the foremost are the collection of standards published and maintained by the International Standards Organisation ("**ISO**"). The need so to do is explicit in the CS2:1995 and CS2:2012 standards. For example, CS2:1995 references ISO:

> "A system of third party certification of the manufacturer to the quality standards of ISO 9002 is designed to ensure [compliance with British Standard 4449:1998] is being carried out." (CS2:1995 page 2, paragraph 8)

and CS2:2012 refers to international standards in general:

> "Review of the CS2 comprises two stages. Stage 1 of the review is to update the technical specification and quality assurance system for steel reinforcing bars **to align with the quality and performance levels as stipulated in the latest international standards**, with due consideration of the conditions and practices of the local industry." (CS2:2012 page 3, "The Construction Standard")

3.3. CS2:1995 states that:

> "The long term objective is to rely on the third party certification of product conformity based on testing and continuous product surveillance and on the quality assurance of stockists. The purchaser would know the quality of the reinforcement being received and would not need to carry out further testing for quality assured reinforcement." (CS2:1995 page 2, paragraph 12)

[1] As explained in my response to Question 2, the testing results generated by MTRCL's HOKLAS accredited laboratory and used in support of the conclusion that a strength reduction factor should be adopted for rebar used in the Project are stated to have been generated in the period from 2010 to 2019. It follows that they would have been obtained under the auspices of both CS2:1995 and CS2:2012.

3.4. This is stating that the long term goal of CS2:1995 is to reach the situation extant in most other countries, where the national standards do not require purchasers' testing provided manufacturers' testing is deemed adequate. This is in line with the ISO over-arching standard ISO 3951-2:2013, which states that standards should allow for a reduction in testing/inspections in order "*…to reduce inspection costs (by means of a switch to a smaller sample size) should consistently good quality be achieved*". What this is saying is that the ISO committee, i.e. the community of experts, advocates a flexible approach for quality assurance when applying standards, rather than dogmatic adherence to rules.

3.5. I have reviewed the rebar testing results [CC11/7252] in relation to the rebar used in the project. In my opinion, the condition of "consistently good quality" referred to by ISO 3951-2:2013 would seem to have been met because: (i) the first testing omission occurred after 5983.350 tonnes of rebar had been tested (in 179 batches); and (ii) all of the tests performed on those batches had the same result (i.e. a pass on all criteria on all tests). However, the comment that this condition "would seem to have been met" is inadequate; we should measure the confidence with which the condition was met.

3.6. We can ask the question: what is the probability that any batch of rebar not tested would have failed, had it actually been tested? We have evidence from the tests that were performed, and we should use that evidence to answer this question. Here we note that, elsewhere in the Verification and Holistic Reports, the reports adopt the approach of assuming a binomial model, in which tests result in either a Pass or a Fail grade. The inadequacy of that approach is highlighted by applying it here: since there have been no failures throughout the project, a binomial analysis would indicate that there is a zero probability of any failure occurring and hence that further testing would have been irrelevant, so the missing tests are irrelevant. That this is an inadequate basis for analysis indicates the dangers of adopting the binomial approach in general, when data is continuous rather than discrete.

3.7. A more appropriate and realistic approach would be to ask "What is the probability that one of the untested samples would have failed, based on the variability in the results available?" All batches tested passed the test criteria stated in CS2:1995 and in CS2:2012. However, if they only passed by a small margin, or if they passed by a large margin but with considerable variation in the results (i.e. the materials were not of a consistent quality), then the probability that an untested batch might have failed could be high. Testing the hypothesis that an untested batch would have passed is a routine calculation in mathematical statistics. This is calculated from the sample estimate of the population mean and variance (see Note 9) [2], and the worst case is less than 0.2% (see Note 1).

3.8. Based on this calculation, the chance that an untested batch would have failed the criteria specified in CS2:1995 and CS2:2012 is less than 2 in 1,000. In comparison, CS2:1995 and CS2:2012 both only require such chance of failure to be 1 in 10 (see Note 2).

3.9. It is therefore my opinion that the level of confidence is 10 times higher than necessary under the standard.

3.10. Calculation of the probability of failure in the untested batches shows that the testing of the rebar after delivery to the site (i.e. the Purchaser's Testing) does meet the principle

[2] "Notes" are set out at the end of this report.

and the intention of CS2:1995 and CS2:2012 and of similar national standards (e.g. BS 4482:2005, AS/NZS 4671:2001) and international standards (i.e. ISO/TC 17/SC 16). Further, the overall proportion specified by CS2:1995 and CS2:2012 (that 3 in every 10,000 should be tested) was maintained, due to rounding (i.e. because the requirement is to test a minimum of 3 specimens (see Note 9), regardless of the size of the batch (0 to 60 tonnes) and because the number of specimens to be tested in larger batches is rounded up).

3.11. However, whilst rigorous analytically, this is a naïve assessment - not the best we can do to answer the question, given the information available. This is because the 'untested' rebar were not, in fact, untested; they actually were tested by the manufacturer. CS2:1995 and CS2:2012 both specify that the Purchaser's test, i.e. the missing test, is not a new test but a verification of the Manufacturer's test. CS2:1995 states:

> "Under a system of quality assurance, the primary responsibility for testing of the reinforcement and ensuring its compliance with the British Standard lie with the manufacturer of the reinforcement" (CS2:1995 page 2, paragraph 8, "Quality Assurance")

3.12. Clearly, since there were no missing manufacturers' tests, the primary means of ensuring compliance was met. Similarly, CS2:2012 states:

> "Under a system of quality assurance, the responsibilities for testing of the steel reinforcing bars and ensuring its compliance with this Construction Standard lie with the manufacturer. A third party certification of the manufacturer's quality management system by an accredited certification body to the requirements of ISO 9001 is required to ensure that the manufacturer's system of quality assurance is being implemented successfully." (CS2:2012 page 3, "Quality Assurance", paragraph 1)

3.13. The relevant question therefore is whether the Purchaser's Testing confirmed the Manufacturer's Testing, i.e. does it "*… ensure that the manufacturer's system of quality assurance is being implemented successfully*" (CS2:2012) and does the purchaser "*…based on testing and continuous product surveillance and on the quality assurance of stockists …… know the quality of the reinforcement being received*" and hence "*would not need to carry out further testing for quality assured reinforcement*" (CS2:1995). This we can test rigorously, by comparing the two sets of results (Purchaser's Test results and Manufacturer's Test results) for all the batches tested twice (i.e. tested by the manufacturer of the rebar and tested by MTRCL's HOKLAS laboratory after it was delivered to site). At the time that the first omission occurred, the probability that the manufacturer's estimate was incorrect, based on the Purchaser's Testing, was less than 5%, and hence the CS2:1995 and CS2:2012 requirement of 95% confidence was satisfied by the testing performed by MTRCL's HOKLAS laboratory on the rebar that had been delivered to the site (see Note 4).

3.14. In summary, the answer to Question 1 is that:

(a) a rigorous statistical analysis shows that the confidence in the rebar that was not tested by MTRCL's HOKLAS laboratory after delivery to the site (i.e. not "re-tested") exceeded that required by both CS2:1995 and CS2-2012; and

(b) even though some batches of rebar were not re-tested after delivery to the site, both the spirit and the intention of the applicable standards were met overall.

Question 2

What are your comments on the statistical methods and analysis referred to in Section 4.3 of the Verification Report for the purposes of assessing the strength of the rebar used in the Project?

3.15. The Verification Report (section 4.3.2) uses data from samples tested during the period 2010 to 2019 to inform decisions about samples not tested during the period 2014 to 2019. The CS2:2012 standard was introduced during this period with a specific claim to increase standards.[3] We would therefore expect, if the new standard met its objectives, that there would be successively less failures reported by MTRCL's HOKLAS accredited laboratory as the new standard took effect. This means that the 55 failures might reasonably have been expected to have disproportionately occurred early in the period of 2010 to 2019, and that, during the period of Contract 1112, a lower proportion of failures might have occurred. Hence, in order to ensure that this analysis was comparing like with like, a test for time-series effects in the data should have been conducted prior to its use for this purpose (see Note 9). There is no reference to the results of such an analysis in the Verification Report.

3.16. The Verification Report applies the results of tests of rebar from all manufacturers without (apparently) allowing for the considerable variance between manufacturers. There has apparently been no attempt to cross-reference the source of the 55 failed samples (at MTRCL's HOKLAS accredited laboratory) with those manufacturers supplying rebar under Contract 1112, and specifically those manufacturers which had supplied rebar that was not subsequently re-tested.

Question 3

What are your comments on the conclusions drawn in Section 4.3 of the Verification Report? In particular, is it reasonable to adopt a strength reduction factor of: (i) 4% for rebar with a diameter equal to or more than 16mm; and (ii) 13% for rebar with a diameter of less than 16mm?

3.17. The analysis is flawed because it assumes that all rebar which was not re-tested should be treated as if it were within the worst 0.05% of all rebar seen by MTRCL's HOKLAS accredited laboratory in the last 9 years. The probability that this could actually have happened is so low that it is effectively impossible. Indeed, it is lower than the estimated probability of picking a specific pre-determined single atom at random from the entirety of the Universe (see Note 5). Closer to home, the probability could be compared with the probability that MTRCL's HOKLAS accredited laboratory has passed as fit for use a batch of rebar which was actually substandard (see Note 6). Hence the logic of the Verification Report, if followed consistently, would be that remedial measures should be required for every structure constructed using rebar tested by MTRCL's HOKLAS accredited laboratory, and to do otherwise would be to hold Leighton to a substantially higher standard of proof of quality than contractors in general.

[3] CS2:2012 was published under the authority of the Standing Committee on Concrete Technology (SCCT) in November 2012.

Question 4

Based on relevant quality assurance standards, how would you assess the adequacy of the testing conducted on the batches of rebar that were used in the Project?

3.18. The adequacy of the testing should be assessed by reference to the relevant quality assurance standards. I have presented this assessment in my answer to Question 1.

3.19. As explained above, the relevant quality assurance standards clearly means CS2:1995 and CS2-2012 but may also be interpreted as including other national and international standards. As stated previously, CS2:1995 states:

> "A system of third party certification of the manufacturer to the quality standards of ISO 9002 is designed to ensure [compliance with British Standard 4449:1998] is being carried out."

And the terms of reference of CS2:2012 were:

> "Review of the CS2 comprises two stages. Stage 1 of the review is to update the technical specification and quality assurance system for steel reinforcing bars **to align with the quality and performance levels as stipulated in the latest international standards**, with due consideration of the conditions and practices of the local industry."

3.20. It is therefore necessary to include, as being relevant, the guidelines given by the ISO and specifically ISO 3951-2:2013, which states that it is acceptable "*…to reduce inspection costs (by means of a switch to a smaller sample size) should consistently good quality be achieved*" (see Note 7).

3.21. As explained in my answer to Question 1, the level of confidence in the rebar that was not tested after it was delivered to site exceeded that required by the CS2:1995 and CS2-2012 standards and, even though some batches of rebar were not re-tested on site, both the spirit and the intention of these standards were met overall.

Conclusion

3.22. It is my opinion that the statistical requirements stated in the CS2:1995 and CS2:2012 standards have been met, both in principle and by reference to rigorous mathematical analysis. Furthermore, it is my opinion that, if Leighton were required to meet a higher standard than that already met by the as-constructed structure, this would be discriminatory, in so far as I have demonstrated that the higher standard being demanded in the Verification Report would, if applied consistently, call into question the safety of all structures that have used materials tested in MTRCL's HOKLAS accredited laboratory.

4. Coupler Connections

4.1. The Verification Report does not involve any sample based testing in respect of the coupler connections at the areas of site relevant to the Extended Inquiry (i.e. NAT, SAT and HHS). Rather, a strength reduction factor of 35% has been adopted for coupler connections at these areas.

Strength reduction factor

4.2. Section 3.2.12 and Table 5 of the Verification Report introduces the strength reduction factor. The relevant row in Table 5 is extracted below:

Table 5: Summary of Findings and Recommendations under Part 1b

Issues	Findings	Recommendations
e) Workmanship of coupler connections	Defective rate of 35% was adopted for coupler connections at NAT, SAT and HHS (Section 4.2.6).	Strength reduction factor of 35% is adopted for coupler connections in Part 2 (Section 4.2.6).

4.3. Section 4.2.6 of the Verification Report states: "*In the absence of any other alternative evidence or data, a strength reduction factor of 35% has been adopted…*"

4.4. This figure of 35% appears to be derived from the analysis of data in Appendix B3 of the Holistic Report [OU5/3229] where the defective rates of 36.6% and 33.2% were derived for EWL and NSL respectively. It appears that the figure of 35% is based on the combined samples of the EWL and NSL and the defective rate is obtained by applying the criteria that: (i) the engagement length of the coupler connections must be 37mm; and (ii) no more than 2 full threads can be exposed out of the coupler.

4.5. I refer to Table 1 and related paragraphs of my report for the Original Inquiry. For the reasons explained in my report, the defective rate of 35% noted in the Holistic Report (i.e. which is for the combined EWL and NSL samples) should be reduced to 9.4%. It follows that the strength reduction factor should also be 9.4%.

Conclusion

4.6. It is my opinion that the calculation of the strength reduction factor is flawed and does not put the available data (obtained from opening up the structure and hence gathered at not inconsiderable risk to the safety of the structure) to best use.

4.7. I have presented an alternative calculation of the strength reduction factor, based on sound statistical principles, which I suggest should be used as a basis for comparison with known values for the spare capacity in the structure in order to assess whether any further work is required to ensure that the structure is safe.

5. Expert Declaration

5.1. I understand that my primary duty in preparing this report and giving evidence is to the Commission of Inquiry, rather than to the party who engaged me and I have complied with that duty.

5.2. I have endeavoured in this report and in my opinions to be accurate and to have covered all relevant issues concerning the matters stated which I have been asked to address.

5.3. I have endeavoured to include in my report those matters, which I have knowledge of or which I have been made aware, that might adversely affect the validity of my opinion.

5.4. I have indicated the sources of all information that I have used.

5.5. I have not, without forming an independent view, included or excluded anything which has been suggested to me by others (in particular my instructing solicitors).

5.6. I understand that:

(a) My report, subject to any corrections before swearing as to its correctness, will form the evidence to be given under oath or affirmation.

(b) I may be cross examined on my report by a cross examiner assisted by an expert.

(c) I am likely to be the subject of public adverse criticism if the COI concludes that I have not taken reasonable care in trying to meet the standards set out above.

5.7. I believe the facts I have stated in this report are true and that the opinions I have expressed are correct.

Barrie Wells

Digitally signed by Barrie Wells
DN: cn=Barrie Wells, o=Conwy Valley Systems Limited, ou=Cowny Valley Systems Limited, email=Barrie.Wells@ConwyValley.com, c=GB
Date: 2019.09.13 09:51:44 +01'00'

BARRIE WELLS

13 SEPTEMBER 2019

Note 1: Calculation of probability that an untested batch would have failed

We can calculate the mean and variance for all tests undertaken and use a Student's t-test to quantify the probability that any other value would lie outside a certain bound. Using the bounds given in CS2_2012, tables 5 and 8, we calculate the probability that an individual specimen would fail and then, assuming independence and using the scenario trees in Note 3, the probability that a sample taken from a batch would fail.

For a test for Yield Stress meeting the minimum requirement, we calculate:

Table 2

Steel Grade	Mean (Tensile Test Yield Stress)	Variance	Probability (Specimen Fails)	Probability (Sample Fails)
460	545.0 MPa	1228.9	0.0174	0.0018
500	570.5 MPa	863.1	0.0147	0.0013

And similarly for the tests on upper bound for yield stress (critical value 650MPa) and for the ratio:

*R*m/*R*e = tensile strength / yield stress

(critical value 1.35)..

Each test results in a different probability of failure; the highest probability provides a conservative estimate; alternative conditions are allowed for a pass and it would be possible to use lower probabilities by choosing which criteria to adopt in each test, but since a conservative estimate exceeds the requirements of CS2:2012 by more than an order of magnitude, there is no need to conduct the analysis to that degree.

Note 2: 90% probability of 95% of the results meeting the criterion

CS2:2012 states:

> "**1.6.1 General**
> The characteristic values as given in Table 5 are (unless otherwise indicated) the lower or upper limit of the statistical tolerance interval at which there is a 90% probability ($1-\alpha = 0.90$) that 95% ($p = 0.95$) or 90% ($p = 0.90$) of the values are at or above the lower limit or at or below the upper limit respectively. This quality level refers to the long-term quality level of production."

The natural interpretation of this statement is that test results should be assessed against a 95% (P = 0.95) point on a z or t test, to calculate a probability of any possible outcome occurring, and that this probability should be required to exceed (or not exceed, depending as the outcome is desired or not) 0.9 (i.e. 90%). However, since all such results calculated herein would pass either a 90% or 95% threshold, the distinction is not relevant.

Note 3: Failure of a batch versus failure of a specimen

CS2:2012 states: "If the requirements have not been met, two additional samples shall be provided from the same batch and additional tests shall be carried out."

(N.B. the CS2:2012 usage of 'samples' here corresponds to 'specimens' – see Note 7)

This means that failure of a batch occurs only if there are two failures out of three specimens. Hence, writing S for success (pass) and F for fail, failure occurs under any of the following conditions (for a batch requiring 3 specimens to be tested):

FF
FSF
FSSFF
FSSFSF
SFF
SFSF
SSFF
SSFSF
SFSSFF
SFSSFSF
FSSFSSFF
FSSFSSFSF
FSSSFSF
FSSSFF

And a batch may pass (Success) from any of the following sequences:

FSSSS
FSSFSSS
FSSFSSFSS
FSSSFSS
SFSSS
SFSSFSS
SSS
SSFSS

Corresponding scenario trees may be constructed for larger batches, requiring 6, 9, 12, 15 or 18 specimens to constitute the sample for the batch (18 being the largest recorded in this project). Hence, given the probability of a single specimen failing, P(F) (with corresponding probability of success, P(S) = 1 – P(F), we may calculate any of the above combinations by multiplying together the independent probabilities. Then we calculate the probability of failure of a batch by adding together the 8 individual possible ways a failure may occur, and check the arithmetic by adding together the 6 possible ways a Success may occur and confirming the two sums sum to unity.

Note 4: Hypothesis Test: Purchaser's test results confirm Manufacturer's test results

The Purchaser's test results are intended to confirm the Manufacturer's test results. Statistically speaking, this means we expect the two samples (sample of steel tested by the manufacturer and sample of steel tested by the purchaser) to have similar average values. To put it another way, there is almost bound to be a difference, because of natural variation from specimen to specimen, but is the overall difference significant? We can test this by calculating the difference between the two means and dividing it by the Standard Error, normalised by the sample sizes. This gives a value which should follow the Student's t distribution:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and hence the hypothesis, that the two samples are from the same population (i.e. that the Purchaser's test results confirm the Manufacturer's test results) can be tested by comparing the calculated t value with the tabulated t statistic.

Adopting a 95% (two-tailed) confidence test, in line with the standards, we should check whether, in at least 90% of the cases, this confidence is exceeded, and hence the CS2:2012 criterion of 90% confidence in a 95% cut-off is achieved.

From the information available it was not always possible to match the actual rebar used to the specific tests on the mill test certificate (hence the classification, under CS2:2012, as Class 2 rather than Class 1). Of the batches where this was possible, > 90% passed the two-tailed 95% test (i.e. t-statistic significance at the 2.5% level).

Note 5: Probability that all untested rebar would have failed

The MTRCL's HOKLAS accredited laboratory recorded 55 field batches out of 110,000 batches tested, yielding a probability of failure, P(F), of 0.0005 for any one batch.

The probability that 2 batches would have failed, assuming independence, is the result of multiplying the probabilities of each batch individually failing. In this case, that is

$$0.0005 \times 0.0005 = 0.00000025$$

Applying this to the 162 batches which were not tested during Contract 1112, the probability that all 162 batches would have failed, assuming independence, is zero followed by a decimal point followed by 534 zeroes before the first non-zero value. By way of comparison, the number of atoms in the solar system is estimated to be zero followed by a decimal point followed by 52 zeroes before the first non-zero value. The number of atoms in the known, observable universe is variously estimated as being a number with 78 to 82 zeroes.

For background on these estimates, refer e.g. to https://www.universetoday.com/36302/atoms-in-the-universe/
They are presented solely for the purpose of giving context to events with incredibly small probabilities and have no independent worth.

Note 6: Probability that a batch of rebar passes CS2:2012 when half of all rebar are below standard

The purpose of this testing is to minimise the chance of using 'bad' materials. Testing cannot establish an absolute answer to the question of whether or not supplied materials are fit for purpose, it can only give an indication. This indication is stated as a probability. The purpose of this testing is to evaluate the likelihood of any number of materials not meeting a specified criterion

In the case of materials consisting of individual specimens, it is generally neither economically feasible nor scientifically necessary to test every specimen for conformance with the standard. In the case of CS2:2012, tests are destructive and hence it is not possible to test every specimen. A sampling procedure is therefore required. CS2:2012 does not give any information on the basis for its sampling procedure, it simply states numbers (3 specimens in every 60 tonnes – say 3 in 10000 rebar).

Clearly, there is the possibility that a sub-standard batch might be passed and there is the possibility that the only three defective rebar in the entire batch are the ones selected for test, so that an entire batch of perfectly good rebar is discarded. Because this is a statistical procedure, these two eventualities (referred to as Errors of Type I and Errors of Type II, 'false positive' and 'false negative') cannot be ruled out. A good testing regime can only hope to minimise their probability.

We will look at the probability of a false positive: passing for use, rebar which is actually substandard. The Verification Report gives no breakdown of specimen results so we will use the results from Contract 1112.

Specimen failure has a probability $P_f = 0.0147$. If, in a batch of 10000 rebar, 3 are non-defective and 99997 are defective, the probability that specifically the 3 non-defective rebar would be selected for testing is

$$(P_f)^3 \text{ x } (1 - P_f)^{99997} \approx 0.5*10^{-66}$$

which is roughly the probability of successfully choosing, at random, a specific, pre-determined atom out of all the atoms in our galaxy. Unlikely as this is, it is substantially more likely than the situation, postulated in the Verification Report, that would give rise to the strength reduction factors suggested.

The corollary is, therefore, that all constructions that have used rebar tested by MTRCL's HOKLAS accredited laboratory should be subjected to remedial measures, for the same standard of safety to be assured.

Note 7: ISO standards and CS2:2012

The terms of reference for the update of CS2:1995, which became CS2:2012, included:

> "Review of the CS2 comprises two stages. Stage 1 of the review is to update the technical specification and quality assurance system for steel reinforcing bars to align with the quality and performance levels as stipulated in the latest international standards, with due consideration of the conditions and practices of the local industry."

It is my opinion that "to align with the quality and performance levels as stipulated in the latest international standards" means the intention and purpose of the new standard includes the over-arching principles of international standards, of which the foremost are the collection of standards published and maintained by the International Standards Organisation (ISO).

ISO standards comprise a hierarchical structure, with families of specific standards grouped under general guidelines. ISO 3951-2:2013 "Sampling procedures for inspection by variables -- Part 2: General specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection of independent quality characteristics" covers general principles for a class of standards that includes the specific group of 'Steels for the reinforcement of concrete', which in turn includes ISO 10144, 15630, 1620, 6935 and others.

In particular, the standard ISO 3951-2:2013 is invoked in Note 8 to establish the over-arching principle that a demonstration of consistent quality should lead to relaxation of testing conditions, in situations within its remit, which includes CS2:2012.

Note 8: ISO standards and the relaxation of inspection criterion when consistently high quality is achieved

In the case of materials consisting of individual specimens, it is generally neither economically feasible nor scientifically necessary to test every specimen for conformance with the standard. In the case of CS2:2012, tests are destructive and hence it is not possible to test every specimen. A sampling procedure is therefore required. CS2:2012 does not give any information on the basis for its sampling procedure, it simply states numbers (3 in every 10000).

Where sampling procedures are described in more detail, e.g. ISO 3951-2:2013 "Sampling procedures for inspection by variables -- Part 2: General specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection of independent quality characteristics", we see that sampling procedures are not intended to be rigid or fixed for the duration:

<<The objectives of the methods laid down in ISO 3951-2:2013 are to ensure that lots of an acceptable quality have a high probability of acceptance and that the probability of not accepting inferior lots is as high as practicable. This is achieved by means of the switching rules, which provide the following:

- automatic protection to the consumer (by means of a switch to tightened inspection or discontinuation of sampling inspection) should a deterioration in quality be detected;

- an incentive (at the discretion of the responsible authority) to reduce inspection costs (by means of a switch to a smaller sample size) should consistently good quality be achieved.>>

ISO are saying that if a process is providing specimens of a consistently high quality, the sampling density may be reduced.

The community of experts, as represented by the International Standards Organisation committees, advocates a flexible approach over dogmatic adherence to rules.

This is not necessarily to say that the SCCT (Standing Committee on Concrete Technology) are incorrect in their specification, nor to imply that their standard CS2:2012 requires modification, but does show that the oversight (of not testing all batches in strict accordance with CS2:2012) met the spirit if not the letter and we can show that it did not significantly affect the results of conformance testing as a whole, as it affects the quality of the completed structure.

Note 9: Terminology

Specimen versus Sample

I have adopted the convention of referring to a single item as a specimen and a collection of specimens as a sample. This distinction is not always adhered to in CS2:2012 nor the Verification Report, which can in some circumstances be confusing. This adopted terminology is standard in statistical texts, although 'specimen' is often replaced with 'observation'. Here, the word 'specimen' is more natural because here it is referring to actual physical specimens, not just observations made on specimens. This terminology is also consistent with the use in the phrase 'random sample', meaning a collection of specimens picked at random.

Sample estimate of the population mean and variance

The actual mean and variance (i.e. measure of the variability) of all specimens in a class (i.e. the 'population') is not usually known, so we estimate it by calculating the mean and variance of the sample for which we do have values. It is important to remember that these are only estimates, not the actual values.

Time Series

A time series is simply a set of values which are measured at particular moments in time. If the data change with time, then comparisons amongst specimens will be affected by when the measurements were taken. Not all data taken over a long period of time will exhibit a trend. For example, lower infant mortality rates and changes to social conditions mean that family sizes are generally smaller now than in the past, although not every family is smaller. On the other hand, the gender ratio in families is unlikely to have changed, so not all statistics measured over time exhibit a trend. Nevertheless, if we want to compare data across time, it should be checked first, as otherwise an unidentified trend can invalidate conclusions we may make.

Note 10: What Is the Purpose of Standards Documents?

The purpose of this type of testing is not to evaluate material based on a specific criteria that states "above X is sound; below X is unsound".

It is true that the relevant standards, CS2:1995 and CS2:2012 ('CS2 Standards') state a value, for each test type, that the specimen must achieve, but it is not the case that, if the specimen achieves this value then the whole batch must all be sound, and neither is it the case that if the test specimen fails, then the batch should be rejected for use.

For the sake of brevity, we will refer to 'strength tests', but the actual CS2 Standards specifies many tests for each specimen.

Looking first at "if the test specimen achieves this value then the whole batch must all be sound", it should be obvious that if, in a batch of 100 of which 97 have a strength that does not reach the standard and only 3 that do, it could happen, just by chance, that those 3 "bad" specimens were selected to be tested (which 3 are selected is just random, because there is unlikely to be any visual indication that any one differs from the others). In that case, an entire batch (of 97) unsound specimens would be accepted, and used in the structure. It is unlikely, but possible. However, even that eventuality would not necessarily make the structure unsound, depending on the amount or margin by which the 97 failed to achieve the standard value.

Looking secondly at " neither is it the case that if the test specimen fails, then the batch should be rejected for use", the CS2 standards explicitly state that, if a specimen is tested and found not to achieve the critical value, then the recommended action is to take two substitutes and test those instead. Clearly, "failure" is not absolute, we get a second bite; not even of the same cherry, but we get to choose a different cherry.

If that is not the purpose, then what is the purpose of tis type of test?

The purpose of this testing is to evaluate the likelihood of any number of specimens meeting, or not meeting, an "ideal" strength criterion, and by how much. Whatever the test results, regardless of whether or not all the specimens achieve the critical value, we should be able to make a probabilistic prediction of the strengths of all specimens, and the engineers can then use those figures to say whether or not the structure would be sound if constructed using those materials. However, we do not do this. Instead, we use a shorthand, based on experience: we say that if a certain number of specimens pass, then everything will be OK. The CS2 standards are therefore just a "rule of thumb", no more. It is certainly possible to be more rigorous, this could be achieved at comparatively little extra effort, but the entire construction industry, from academics and engineers setting standards through politicians and legislators specifying and implementing those standards and on to engineers and craftspeople following the standards, all agree that approximations are acceptable, a simple rule of thumb can replace rigorous testing. This is not disputed, it is universally accepted and followed, and forms the basis for the construction industry worldwide.

Furthermore, the standards experts accept that testing should not be dogmatic; if the same material sources are used throughout a construction project, then testing should be allowed to become less strict as confidence is built up. This is a recognition of the obvious observation that, if no quality problems have been encountered to date, and nothing has changed in the process, then the chances of a quality problem occurring in the future are lower than they were at the start; and, further, that the longer the project continues, the lower those chances become, and so the less critical the testing becomes. Specifically, it is recommended, by the International Standards Organisation Committees, that we should allow for less testing, as time progresses with no problems being encountered.