July 17, 2012

Standards Australia
Level 10
The Exchange Centre
20 Bridge Street
Sydney

LANCASTER UNIVERSITY
Dr. David Lucy
Department of Mathematics & Statistics
Lancaster University
LA1 4YF
United Kingdom

+44 01524 594338
d.lucy@lancaster.ac.uk
http://www.maths.lancs.ac.uk/~lucy/

Ref: AS5388 Forensic Analysis
5388.3 Part 3: Interpretation

To whom it may concern,

Firstly, the Australian Government is to be congratulated for considering forensic evidence interpretation and evaluation as an integral, but separate part of forensic science. As far as I am aware the Australian Government is the first to do so, and I hope that this initiative will help to clear up the problems and inconsistencies seen with forensic evidence interpretation in other jurisdictions.

As these standards are to some extent pioneering it is vital that they are informed by the scientific work undertaken over the last thirty years or so on forensic evidence evaluation. This is because scientific standards have an influence on practice and policy, both into the future, and internationally, as other governments look towards these standards to inform their own guidelines. To this end it is important that confusion, inconsistencies and logical fallacies are avoided in the final form of these standards.

I wish to comment on a number of points, each of which is a response to a point in document AS 5388.3201X. **The comments here will therefore be numbered, the numbering referring to the respective point in AS 5388.3201X.**

4.1 General - AS 5388.3201X says:

"The default position is known as the null hypothesis, which would usually be that

1. *the action could not produce the outcome;*
2. *X is not a particular thing; or*
3. *items do not share a common origin.*

The examiner should follow an approach that seeks to evaluate the null hypothesis."

A general idea of a *null hypothesis* can indeed be any of the three selected senses of what a *null hypothesis* can be, but it should not be true that an examiner should seek to evaluate it. This ap-

proach is part of the frequentist paradigm, and has little relevance to modern forensic evidence evaluation. This approach can in fact be wholly misleading in terms of evidence evaluation.

This is because the *null hypothesis* is not directly relevant to the question being asked of the evidence. The question of forensic relevance is that relating to the offence, not a question relating to **not** the offence. By simply looking at the *null hypothesis* in isolation the finders of fact are simply running around in logical circles.

For instance, were one to make the observation that a fragment of glass associated with some offence contained 0.2% iron. A similar glass fragment, also with a 0.2% iron content had been recovered from some suspect, and that the probability of observing a similar fragment of glass from all possible glass fragments also with an iron content of 0.2% were 1 in 10,000.

Does this mean that there is a 1:10,000 probability that the two pieces of glass are not from the same source? Do we go on to say that therefore there is a 99.99% probability that these two fragments are from the same source? No, neither of these two evaluations would have any foundation in either the theory, or practice of statistical evidence evaluation.

The nature of realistic continuous observations entail that the probability of any state, even that state which is most probable, occurs with a low probability, simply by virtue of it being continuous. A low probability in this case indicates very little on its own.

This is why a focus on probabilities associated solely with *null hypotheses* are of little use when evaluating forensic evidence.

4.4 Limitations - AS 5388.3201X says:

"In qualitative examinations, variation exists but is not usually susceptible to quantitative analysis that yields an uncertainty value."

There are very few areas of forensic science not amenable in some way, at least in principle, to quantification. It may be better for those for which quantification is in principle impossible, should be considered as evidence of "opinion", and thought of as a different sort of evidence than the more usual parts of forensic evidence.

5.1 General - AS 5388.3201X says:

"(c) Consider the probability of error.

Is "error" being used in the statistical sense, or in the sense of "making a wrong decision". If the former a probability cannot exist for a measure of uncertainty because it itself is a measure of uncertainty, and would be a tautology. If the latter, then, except in a very few trivial cases, cannot yet be calculated. This would place most forensic science beyond what is currently

possible.

5.2 Statistical analysis - AS 5388.3201X says:

"If data cannot be applied to a suitable statistical distribution or transformed to fit a suitable distribution, the use of appropriate non-parametric statistical descriptors should be considered. In such circumstances any information derived should be treated as an opinion"

There are many methods of handling observations which are not distributed in any simply parametrised way, all but a few degenerative cases, which are not encountered in forensic science, can be treated using methods such as log-concave estimation, or kernel density estimation. For some continuously distributed observation it is common to evaluate a likelihood ratio using between item distribution based kernel on density methods, although for some a normal assumption is possible. Is the former to be regarded as "opinion", and the latter "fact"? The standard would make no sense whatsoever in a statistical sense were this principle included.

Where results from a sample of the population are extrapolated to cover the population, the statistical approach used should be documented. The confidence levels, or credible intervals, for the extrapolation shall be calculated and recorded.

Why are credible intervals, or confidence intervals so important to the process? These are merely summaries used by statisticians for quick comparisons. It is the distributions which are important for probability calculations, the summary measures are useful in that sometimes they are sufficient to characterise the distribution, but the estimates of the probabilities are the important features.

5.3 Comparative testing - AS 5388.3201X says:

When conducting comparative testing, the question being considered is whether or not the items being compared have a common origin. If a common origin is supported, any differences observed shall be capable of being adequately explained.

This contradicts the usual meaning of *null hypothesis,* and that given in Section 4.1 of AS 5388.3201X.

8.2 Estimating probability - AS 5388.3201X says:

"Where suitably representative and statistically valid databases exist from which reproducible numerical values may be obtained, a numerical probability value may be calculated in support of the opinion. This may be expressed as a likelihood ratio or as a frequency."

Likelihood ratios are the current focus of research and development for forensic evidence evaluation. A frequency implies a completely different approach, and one which is at odds with a likelihood ratio approach. The Australian Standards should have a single clear position, and not mix and match incompatible measures of evidential value.

The use of numerical likelihood ratios is well accepted in DNA typing. Many jurisdictions

report the ratio as a result and then convert that ratio into a statement about the weight of the evidence.

This is the currently favoured approach, and in my opinion should be that adopted in so far as it is possible by the Australian standards.

This may again be expressed as a likelihood ratio or as a frequency and used in a qualitative way, for example in a statement such as X (evidence) is more likely if Y (action or person) rather than Z (action or person) or The two items have a common source. The statement represents the level of opinion (expressed verbally rather that numerically) as to the significance of the results obtained.

This statement mixes two forms of statement. A likelihood ratio is about a quantification of evidence, and is the legitimate province of the forensic scientist. However, the statement *"The two items have a common source"* is a conclusion, and is that which should be informed by the evidence, and is a matter for the fact finders (Court, Judges etc). It is not for the forensic scientist to encroach upon the role of others in the legal process.

Summary

This is a brave effort on the part of the Australian Government to formalise the process of forensic evidence evaluation. However, in these early stages the standard lacks coherence, seemingly adopting mutually incompatible paradigms of probabilistic reasoning.

Much of the tenor of these standards seem to be unable to separate conclusions from evidence, would fatally undermine efforts on the part of the forensic scientist following the standards to evaluate their evidence in any rigorous manner.

The field of forensic evidence interpretation and evaluation has come a long way in the last few decades, and it is this scholarship upon which the Australian Government should look to for formulate its standards.

Yours sincerely,

David Lucy