In Australia the jury decides: the reliability and validity of expert evidence: a perspective from forensic pathology

Stephen Cordner\textsuperscript{a,b} and Noel Woodford\textsuperscript{a,b}

\textsuperscript{a}Victorian Institute of Forensic Medicine, Melbourne, Australia; \textsuperscript{b}Department of Forensic Medicine, Monash University, Melbourne, Australia

\textbf{ABSTRACT}

The Innocence Project has enabled, amongst many other things, quantification of contributors to wrongful convictions in the United States. There, individual cases of wrongful conviction have multiple causes, including problems with forensic expert evidence. The Criminal Cases Review Commission in England and Wales has found that amongst unsafe convictions there, forensic expert evidence has been a feature. Apply the English experience with unsafe convictions to Australia, and there would be many more cases than we know of. The Victorian Supreme and County Court Practice Note on Expert Evidence (2014) has reduced the risk that providers of expert evidence could contribute to wrongful convictions, but more needs to be done. The medical sector generally has a huge research capability and capacity. In comparison the forensic pathology sector is tiny and is characterized more by its focus on managing workloads. Conclusions about the cause of death are not testable, and other opinions provided by forensic pathologists are derived from medical sciences, and not from specific evidence-based research. This does not mean that forensic pathology evidence is inherently unreliable, but it does mean that more work is required to better establish that reliability.

\textbf{ARTICLE HISTORY}

Received 4 December 2019
Accepted 28 January 2020

\textbf{KEYWORDS}

Expert evidence; wrongful convictions; validity and reliability; forensic pathology

1. Introduction

What is the right starting point here? We appear in the courts and provide medical evidence there. It seems reasonable to claim a stake in discussions about how our evidence is heard. We have backgrounds talking with patients – and in our work talking with bereaved families – and know how difficult it is for people, especially when under stress, to understand what others say. Furthermore, those others – doctors, scientists, lawyers, police – vary in their communication skills. The significance of this and other imperfections in our criminal investigation, prosecution and trial process has only really become apparent in the light of the US Innocence Project.

1.1. The US Innocence Project

The US Innocence Project has opened eyes in that country to the fallibility of the American justice system, especially with regard to the evidence provided in the criminal trial, its
presentation and evaluation. Based on the project’s analysis of the Innocence cases, it
believes there are ‘staggering’ numbers of wrongly convicted (in this paper that means
factually innocent) people in American jails. The reasons for this conclusion are:

- All the innocence cases are those where biological evidence was involved, and that
evidence was retained and available for DNA profiling;
- Such cases thus tend to be the serious crimes of rape and murder;
- Analysis of these cases shows that the failures leading to conviction of the factually
innocent, wrongly accused person were: mis-identification of the accused, over-
zealous prosecutors, inadequate defence representation, mistaken forensic evi-
dence, lying witnesses of all types including police and experts;
- There is no reason to suppose that these failures would be confined to the class of
cases where biological evidence exists;
- Therefore these failures exist across the board, infecting the criminal trial process in
that country generally;
- Thus there are staggering numbers of wrongly convicted people in American jails –
an entirely reasonable inference you might think in a country where there were
2.3 million people in jail at the end of 2016 (655/100,000 adults). In Australia in
June 2018 there were (on average per day) 42,974 adults in Australian prisons
(221 per 100,000 adults).

What is more, about half the Innocence Project cases have problematic forensic evidence
included as one of the multiple issues related to the mistaken nature of individual convictions.

1.2. The UK Criminal Cases Review Commission and implications for Australia

The UK is one of the only countries in the world with an independent, government funded
agency to handle post-conviction reviews. The focus of the Criminal Cases Review
Commission (CCRC) in England and Wales is the safety of the conviction, not innocence.

Between 1997 and August 2019, the CCRC referred 657 cases to the Court of Appeal
which have been heard. In 441 of these the Court upheld the appeal, dismissing the
appeal in 203 cases, the remaining 13 either being discontinued or the judgement was yet
to be delivered. Are we in Australia similar enough to the UK to apply those numbers
proportionately here? Pro rata, taking into account both national and prison population
differences, if the British experience had applied in Australia, since 1997, there would have
been approximately:

- 297 cases (appeals) referred to the states’ appeal courts
- 200 appeals allowed (that is, around 9–10 per year)
- 92 appeals dismissed (with 5 awaiting judgement)

(The Australian figures were calculated as follows:
Imprisonment rate per 100,000 population Australia: 170 X Population of Australia
X CCRC numbers, divided by
Imprisonment rate per 100,000 population UK: 141 X Population of UK
The estimated population in 2019 of the UK was 67.5 million, and that of Australia was 25.2 million\(^6\). The assumptions in making this estimate were that the proportion of unsafe convictions was approximately the same in each country, and therefore that the number of unsafe convictions would increase as the numbers in prison increased; and that the rate of imprisonment in England and Wales was a surrogate for the rate in the UK as a whole.

These numbers are clearly a little speculative, but based as they are on a system which resembles ours, they should discourage any local tendency to complacency. The only objective comparator we know of is one researched by Dioso-Villa: 71 people who over a period of 93 years between 1922–2015 she regarded as factually innocent and wrongly convicted in Australia\(^7\).

The systemic problems referred to by the CCRC include:

- Insufficient or misguided investigation
- Fabricated or suppressed evidence
- Misconceived expert evidence
- Confessions obtained through duress

The commonest basis of referral to the Court of Appeal by the CCRC remains non-disclosure of material evidence to the defence or to the court at the trial.

‘But failures, miscarriages and cover-ups will always be with us because the ingredients which give rise to them are ever present: pressure to get results, institutional complacency, the temptation to cover up rather than come clean, professional short comings, moral weakness, fear of failure and fear of exposure’\(^8\).

It would be a brave person who would say this comment does not apply to Australia.

1.3. The resurrection of the AAFS (Victorian Chapter) and the Supreme and County Court Practice Note (2014)

The Victorian Chapter of the AAFS was resurrected in the wake of the wrong conviction of Farah Jama. This injustice was initiated by the collection of biological evidence from a contaminated environment. The disaster occurred despite opportunities which arose throughout the investigation, pre-trial and trial processes to realize the truth of the matter. Justice Vincent’s review of the conviction made a number of recommendations, and both it and Julie Szego’s ‘The Tainted Trial of Farah Jama’ are mandatory reading for those interested in how a wrong conviction can occur\(^9\),\(^10\).

The need for a forum in Victoria where leaders from the different components of the criminal justice system could mix to reduce the impact of silo’d thinking was one of Justice Vincent’s recommendations; thus the resurrection of the Society. And before Jama’s case was the miscarriage of justice in R v Klamo – a case squarely in the zone of expert evidence and its reception during a trial\(^11\),\(^12\).

Klamo’s case led to the Supreme Court’s Practice Note No 2 of 2014 governing both the content and procedure to enable adequate pre-trial scrutiny of important expert evidence\(^13\). In preparation for the introduction of the Practice Note there was a large amount of work undertaken at the Victoria Police Forensic Science Department to re-engineer their forensic science reports to better enable the courts to evaluate the scientific basis of their results and conclusions.
When the Practice Note is activated by either the defence or prosecution, the forensic pathologist or physician is obliged to deal in detail prior to the trial with the issues raised. The idea was that as a result of such work, the forensic pathology or medicine issues in dispute could be resolved or at least distilled to their essence prior to the trial. This would reduce to the necessary minimum the burden of expert forensic pathology or medicine evidence within the actual trial. The Practice Note has the additional benefit of the expert’s attention being drawn to the important issues prior to the trial, resulting you might think in a stronger expert evaluation of the issue, rather than an off-the-cuff response when the issue arises out of the blue during the trial. The days are long gone we believe when the autopsy or clinical forensic medicine report can be expected to deal, without notice, with all the possible significant issues that might arise during the trial. There has to be an easy way for the further identification of, work on and exploration of these issues before the trial. That cannot be too hard, surely, in 2019.

This does require adversarial thinking to shift. ‘Why should we disclose beforehand what issues we have with the forensic medical evidence?’ the defence might ask. The answer of course is: ‘so that the chances are improved that the evidence actually given is reliable’. Unreliable evidence will not necessarily work in favour of the accused; and the checks and balances of the trial process may not be at their strongest dealing with difficult, problematic, faulty, flawed or unreliable expert evidence. As expert reports are supposed to be shared prior to the trial, there should be no issue of principle with the Practice Note. We applaud the advent of the Practice Note, a step which reduces our exposure, and that of forensic science also in our view, to the possibility of contributing to a wrong conviction. It should be activated more often.

As should be obvious, in forensic medicine we are very concerned not to be associated with a wrong conviction. By the time a case of homicide or serious assault (with the exception perhaps of sexual assault) gets to court, the prospect of conviction is so high we think the following is a matter for serious consideration: the role of the state’s forensic medicine provider, consistent with the fullest possible forensic medicine investigation, truth and transparency, is to reduce to a minimum its potential contribution to a wrong conviction. (This is in contrast to a more general intention to contribute as well as possible to the justice system). This is the forensic medical corollary of the Hippocratic dictum: first do no harm. Adoption of such a principle would have consequences for the way we operate, for our pre-trial involvement, for our knowledge of the brief, and has been previously explored.

The AAFS (Victorian Branch) has decided that the provision and reception of expert evidence is its core interest. The issue of foundational validity, validity as applied to specific issues in particular cases, and reliability are key issues. While simply named they are complicated: they do not apply to all forensic sciences and branches of forensic medicine in the same way or to the same degree.

So, all that is by way of introduction. How should issues of validity and reliability in forensic pathology evidence be managed? We see our contribution to this summit as beginning to unwrap what we as forensic pathologists do when we give evidence. It is only in recent times that we have been encouraged to do this in more detail. As we said above there is no one size fits all approach. Forensic pathology is different to clinical forensic medicine, forensic anthropology, forensic toxicology, forensic molecular biology (DNA), and all these are different to each other in the detail of how the validity and
reliability of their very many elements will be assessed – and while there will be commonalities with the various forensic sciences at the Victoria Police Forensic Services Department, there will be numerous differences too.

2. The forensic medicine and health sectors contrasted

The doctors in the health service sector generally are well educated in the validity and reliability of what they do. At the same time, they are aware of their limits and of what remains to be discovered and learnt. Behind the health service sector are the large sectors of tertiary and quaternary biomedical education and research – indeed the largest teaching and research component of our universities. (In 2018, for example, Monash University had total revenue of $A2.64 B, over $A600 M of which was from the Faculty of Medicine, Nursing and Health Sciences). These public sectors are supplemented by large private sector investment in, for example, pharmaceutical research. These sectors supply ever more informed and competent professionals using new discoveries to tackle cancer, infectious and degenerative disease and other modern scourges.

Lawyers and perhaps courts on the whole are also not primarily concerned with questioning the validity of what is presented as independent impartial evidence from forensic pathology. At this point the contrast with the biomedical education and research sectors becomes stark. Tiny numbers from those huge sectors go into forensic medicine and pathology, into their small institutions scattered sparsely around the country, with little or no research budgets and few teaching responsibilities. There will often be no academic culture of questioning in these small settings; the dominant driver will be managing workload. Research capacity and capability are thin on the ground. The evidence base for demonstrating validity and reliability is small.

The concept of evidence-based medicine (EBM) arose in the early 1990’s in Canada as a reaction to an over reliance by clinicians on their judgement and experience in making treatment decisions. Proponents of EBM refer to hierarchies of evidence, with randomized controlled trials (RCT), meta-analyses and systematic reviews of such trials being regarded as the strongest. The control in RCT refers to controlling for potential biases or confounders. At the bottom of such hierarchies of evidence, regarded as the weakest evidence, sit observational studies such as case reports, case series and formal consensus.

Yes, much of modern medicine relies on random double-blind control trials to, for example, conclude that a particular remedy is more efficacious than another. But some have claimed that medical knowledge established observationally exceeds the knowledge that has come from randomized controlled trials. The contribution of antibiotics generally to human welfare was not the result of massive trials – although the relative benefits of one or other antibiotic in particular circumstances probably have; aspirin for mild headache; appendicectomy for acute appendicitis; the Heimlich manoeuvre for dislodging a food bolus in the upper trachea; adrenaline in the treatment of anaphylaxis. These are all examples of an advance based on observation alone. So observational studies can be affected by biases and confounders, but they can also deliver reliable results.

For obvious ethical and practical reasons, controlled trials are not possible in forensic pathology. The cause of death is perhaps the single main outcome we provide yet is
largely an untestable opinion or conclusion. The difficulties of characterizing the elements of ‘shaken baby syndrome’ using animal and other models are well known.

3. The current framework for expert evidence in forensic pathology

It is interesting to us that when we provide reports for the defence – or for commissions of inquiry – we are presented with all the information that might bear upon the opinion, and much other information which does not. In short, we are provided with the whole brief. This contrasts sharply with our usual role when we write the report relied upon by the prosecution in a trial. We have the information available at the time of our involvement, usually in the very early days (around the time of the autopsy, or the clinical examination in the case of the physician), almost always before the formal interviews of those involved, and thus information which is vulnerable to being wrong.

With this information forming the context for our autopsy findings we come to conclusions about the cause of death and, effectively, the manner of death. Variability is what then characterizes the extent to which further information is sought or provided, conclusions clarified or further questions pursued and answered. The Practice Note is focused on the big issues which can arise: the cause of death and its nature. But it could be applied to the myriad of ‘smaller’ issues, which in particular cases and depending on the circumstances can be of great significance. There is little in the way of direct evidence upon which to base the answers to them. Such questions, which often arise in relation to corroboration, include:

- In a person who has been strangled, how long does it take to compress the neck in order to (a) lose consciousness and (b) die?
- As a surrogate of the above, how long until petechial haemorrhages appear in neck compression?
- What is the force required to cause bruises in the deep muscles of the neck and laryngo-hyoid fractures in cases of strangulation?
- Deaths of infants in cars on warm/hot days – the temperatures involved, how long does it take to die related to age and other factors (as potential indicators in particular cases of the negligence involved)?
- How long might it take for a child to drown in a bath? i.e. What is the minimum time they were unattended?
- How long does it take to die after a .22 gunshot wound to the frontal lobes of the brain? After incising, or partially incising, the carotid artery? After a single stab wound to the heart?
- What does having a stab wound to the chest feel like?
- What activity is possible following a fatal stab wound to the heart?
- How much force does it take to produce subdural haemorrhage in an infant?
- Would not the assailant in this particular case almost certainly have blood on them?

There are almost as many questions as there are cases. When we are asked them in a trial, usually without warning, we respond unprepared. We do not know the potential significance of the question. In practice, most such questions are answered from first principles of the sciences of anatomy and physiology informed variably by accumulated
knowledge and experience. And at present we don’t have a generally accepted way of conveying levels of certainty with respect to our findings and the conclusions drawn from them (given the particular case context), although Pollanen has proposed a hierarchy which would give an indication as to the level of evidential strength\textsuperscript{17}.

One of us, in the interstices of the time available over the last 12 months, has spent the equivalent of about 4 weeks full time, with 3 co-authors, undertaking a systematic review for publication of case reports of self-strangulation by ligature\textsuperscript{18}. Systematic reviews of the literature to produce evidence-based answers to each of the above questions could take similar effort.

4. Validity and reliability

There are slightly different conceptions of validity and reliability in the law cases and in medicine and science. We understand validity to be the quality of being well founded or sound. In relation to a measure or result it means the extent to which the measure or result reflects the truth of the phenomenon. In ISO/IEC 17025:2017 (General requirements for the competence of testing and calibration laboratories), validity means verification by the provision of objective evidence that a given item fulfils specified requirements, where the specified requirements are adequate for the intended use. Validation is the process by which validity is established. Reliability is a term used in Quality Management, but not so frequently in the ISO/IEC standards. We understand reliability to be the quality of being trustworthy or safe. The reliability of a scientific test result or a medical opinion is its stability when arrived at by different observers in different places at different times. Juran, one of the founding fathers of Quality Management defines reliability as ‘the probability that a product will carry out its intended function under specified conditions and for a specified length of time’.\textsuperscript{19} The American Society of Quality views reliability as ‘the probability that a product, system or service will perform its intended function adequately for a specified period of time, or will operate in a defined environment without failure’.\textsuperscript{20}

The Victorian Court of Appeal said in R v Tuite that the touchstone of reliability is proof of scientific validation\textsuperscript{21}. Clearly, formal scientific validation must mean that reliability (of that which has been validated) has been established. Scientific validation – if that means a formal experimental process – in forensic medicine will be very much the exception and not the rule such is the nature of medicine generally and forensic pathology and medicine in particular. In R v Tuite, the Court referred to Daubert’s case from the US Supreme Court and the judge’s gatekeeper role in that country in relation to the assessment of validity and reliability.

In Daubert, the Court said:

\ldots\ The subject of an expert’s testimony must be ‘scientific … knowledge’. The adjective scientific implies a grounding in the methods and procedures of science. Similarly, the word knowledge connotes more than subjective belief or unsupported speculation. The term applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds. Of course, it would be unreasonable to conclude that the subject of a scientific testimony must be known to a certainty: arguably there are no certainties in science.

But in order to qualify as scientific knowledge an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation – i.e.
‘good grounds’ based on what is known. In short, the requirement that an expert’s testimony pertain to ‘scientific knowledge’ establishes a standard of evidentiary reliability.

The overarching subject of the admissibility inquiry, the US Supreme Court said, should be ‘the scientific validity – and thus the evidentiary relevance and reliability – of the principles that underlie the proposed submission. The focus must be solely on the principles and methodology, not on the conclusions they generate.’

The admissibility inquiry must be a flexible one including the following considerations:

• “Whether the theory or technique can be or has been tested
• Whether the theory or technique has been subjected to peer review and publications; and
• The known or potential rate of error, and the existence and maintenance of standards controlling the technique’s operation.”

And in particular cases there might still be a role for a test of general acceptance.

There is reference above, and in common discourse, to ‘the scientific process’. This is not, contrary to what is often thought, a precisely ordained experimental method for all situations. The US National Academies of Science refer to six guiding principles for scientific discovery:

1. Pose significant questions that can be investigated empirically (nb the quality of the question posed is crucial)
2. Link research to relevant theory
3. Use methods that permit direct investigation(s) of the questions
4. Provide a coherent and explicit chain of reasoning
5. Replicate and generalize across studies
6. Disclose research to encourage professional scrutiny and critique

The UK has also been wrestling with the assessment of validity and reliability in expert evidence. There, too, it is the Judge’s role to act as gatekeeper. In 2014, the Lord Chief Justice in the UK published a new Practice Direction on Expert Evidence in Criminal Proceedings which states inter alia:

“Therefore factors which the court may take into account in determining the reliability of expert opinion, and especially of expert scientific opinion, include:

(a) the extent and quality of the data on which the expert’s opinion is based, and the validity of the methods by which they were obtained;
(b) if the expert’s opinion relies on an inference from any findings, whether the opinion properly explains how safe or unsafe the inference is (whether by reference to statistical significance or in other appropriate terms);
(c) if the expert’s opinion relies on the results of the use of any method (for instance, a test, measurement or survey), whether the opinion takes proper account of matters, such as the degree of precision or margin of uncertainty, affecting the accuracy or reliability of those results;
(d) the extent to which any material upon which the expert’s opinion is based has been reviewed by others with relevant expertise (for instance, in peer-reviewed publications), and the views of those others on that material;
(e) the extent to which the expert’s opinion is based on material falling outside the expert’s own field of expertise;
(f) the completeness of the information which was available to the expert, and whether the expert took account of all relevant information in arriving at the opinion (including information as to the context of any facts to which the opinion relates);
(g) if there is a range of expert opinion on the matter in question, where in the range the expert’s own opinion lies and whether the expert’s preference has been properly explained; and
(h) whether the expert’s methods followed established practice in the field and, if they did not, whether the reason for the divergence has been properly explained. 25

Our reading of the above invites consideration of the distinction between science and medicine. As in the above, an analysis of expert evidence under Daubert also allows for forms of expert evidence to be reliably founded elsewhere than in experimental scientific validation. The weight accorded to such evidence might, we suppose, be different.

Neither of the authors is schooled in the philosophy of science or medicine. That says something in itself. We imagine not many judges are either. Stepping into this space requires some engagement with such basic concepts as:

(1) Schools of thought:
   ● empiricism (knowledge obtained by observation or experimentation);
   ● rationalism (opinions or actions based on reason or knowledge and not religious belief or emotional response);
   ● scepticism (impartiality prior to investigation).

(2) Forms of logical reasoning:
   ● Deductive reasoning (starting with a general statement and making observations to reach a specific conclusion)
   ● Inductive reasoning (going from the specific to the general)
   ● Abductive reasoning (starts with a set of observations and proceeds to the likeliest possible explanation eg a jury arriving at a verdict; a doctor arriving at a diagnosis)

In relation to forensic pathology and medicine case work, there are aspects on top of these:

   ● The training, knowledge and competence of practitioners
   ● The involvement of the practitioners in monitored, regular, relevant competency testing, the results of which should be available
   ● The sophistication of the quality management system in the practitioner’s institution, and the engagement with this by the practitioner
   ● Implicit in the above, the existence of meaningful peer review of the scientist’s/doctor’s work
It might be that arguments about reliability in forensic pathology overlap with arguments about reliability in history. What we actually do most of the time is apply observations and medical conclusions, some of which may have subjective elements, to events or possible events that happened in the past. The aim is to more accurately recreate those events. Would it be right to say there is no reliable view of history because that view has not been scientifically validated?

What about differences in expert forensic pathology opinion; what does that say about reliability? Applying the definition used above of reliability, that means either the evidence is not reliable (because reliability means that two experts in different places and time would come to the same conclusion on the available material) or that one witness is seriously mistaken or worse. At the very least, the court may have to distinguish between the reliability of the core aspect of what is being discussed, and the reliability of that knowledge being applied to a particular event or possible event to establish something of relevance.

To understand potential biases introduced by possibly irrelevant information, sequential unmasking might work in some forensic pathology situations. The aim is to make explicit the reliance being placed on circumstantial information in coming to particular conclusions. To take an example: perhaps the autopsy of an infant should be undertaken with no information whatsoever – only that the baby died. The pathologist undertaking the autopsy would not know

- if the baby died suddenly or over a period;
- whether s/he died at home, in a car or in hospital;
- who was present or in the vicinity;
- if it was a car accident, a fall or the injuries occurred at 2am in the home when it was dad’s turn to deal with the crying baby; or
- that an(other) infant has previously been removed from the home.

This type of exercise would be very resource intensive. It would also run the risk of things being missed, as observations and approaches to address what appear to be significant questions in the circumstances of the particular case might not be made or undertaken because those circumstances were not allowed to be known by the pathologist. We do not rule out such approaches as having a contribution to producing more context-free conclusions from forensic pathology in some subsets of cases. We suspect the outcomes will be additional to, rather than replacing, current approaches. This is one area where research would be valuable.

The issue is, in part, in the management of forensic pathology. Hospitals exist mainly to apply and implement knowledge of medicine and a whole range of sciences – and on the whole the application of that knowledge works well. In particular cases the science and medicine as applied to individual patients fails. Our best hospitals, it is no coincidence, also happen to be those hospitals which contribute most strongly to teaching and research. Teaching and research are activities which make sure practitioners keep current, thinking and talking about what they do. So important is all of this to us as a community that we have medical schools which are by far the largest and wealthiest faculties in their universities.
So it should be with our forensic pathology and clinical forensic medicine – and dare we say, with our forensic science – institutions. They should be service, teaching and research centres, supported by meaningful research budgets, working to deal with case-loads at the same time as improving and strengthening the knowledge base of the disciplines validly and reliably.

**Acknowledgments**

We are indebted to Alastair Ross for permission to distil elements of his Forensic Science subject in the Monash University Masters of Forensic Medicine course.

**Disclosure statement**

No conflict of interest, potential or actual, was reported by the authors.

**References**

2. Please note: the rates here are per 100,000 adults; later in this section rates are referred to, and are rates per 100,000 total population. [cited 2019 Oct 20]. Available from: sentencingcouncil.vic.gov.au/statistics/sentencing-statistics/australias-imprisonment-rates
15. Personal Communication. Finance Department, Faculty of Medicine, Nursing and Health Sciences, Monash University.


