

NATIONAL JUDICIAL COLLEGE OF AUSTRALIA
SCIENCE, EXPERTS AND THE COURTS CONFERENCE

Saturday 4 November 2006, Sydney

HINDSIGHT BIAS IN THE PREPARATION OF EXPERT REPORTS

Dr Thomas Hugh
(Emeritus Consultant Surgeon, St Vincent's Hospital, Sydney and
a Director of the Medical Defence Organisation, MDA National)

(This is an edited transcript of Dr Hugh's presentation)

I'm very conscious that the legal literature is large on the subject of hindsight, particularly in relation to factual evidence, but there seems to have been less focus on hindsight bias in expert reports. That is what I am going to address this afternoon. I speak about hindsight bias in medical expert evidence particularly because that's where I have experience, but I think a lot of what I have to say applies to experts in other fields also.

The concept of hindsight bias is pretty simple. It the adverse effect of experts convening to look back down one pathway in a chain of events. The person, on the day the doctor or the nurse in the case of medical adverse event, sees a whole lot of different pathways only one of which will lead to an adverse event. The essence of hindsight bias is what we call medically the retro- spectroscope.

It doesn't equal foresight. Is hindsight a cognitive error ? I'm daring to look at psychology, but in fact, I think the evidence is strong that it isn't. Let's just consider for a moment in this context how vision works.



Various people might see two people looking at each other but if I said to you that is actually a marble urn, you would also see that. If you turn your mind to it, you can make yourself see either one of those two pictures. That tells us that vision is a construct, it's not a photograph.



Here is a series of blobs. The mind's eye is able to link the dots and make that into a pussy cat.

It's the same with our perception of the world around us. Our perception is a construct and a vision is a construct. So is our perception. Our memory of long distance things is, like our memory of what happened a nana-second ago, a reconstruction not a reproduction. In fact, hindsight is a very important adaptive process in human thinking that allows us to make coherent sense of an environment around us that is constantly changing. In other words we can feed into it new information about the outcome of events and make a different construct.

So it is hard-wired into our brains to utilize hindsight to reconstruct a past event. In fact, if we look at these two groups, the experts and the people involved in the adverse event, the key difference between them is that the 'experts' know the outcome, the people involved in the adverse event did not. So there is an inevitable tendency for the 'experts' to inject the present reality (what we now know) into the reality of the event (what was then known). That is the essential element in hindsight bias.

Hindsight bias also involves counter factorials. A counter factorial is a statement that refers to things that did not happen. If we take a health care worker, a nurse or a doctor and think of the pathway of events as a kind of tunnel that they enter which has forks on the way, some of those forks will lead to a safe outcome and some will not. The person enters that tunnel. The expert will say "Hmm, why didn't she zag or zig?" The experts are considering counter factorials. They are product of hindsight. You can tell when hindsight is operating in an expert report by this tell tale phrase:

"there a number of psychological studies that show it should have been obvious"

It is the giveaway.

What effects does hindsight bias have? It tends to simplify or trivialise the decisions that the person at the sharp end on the day was making. As I have said, it injects the present reality (what we now know) into the reality of the incident. A number of studies show that if a group of experts are given a range of scenarios, their judgements are harsher if the adverse event has a severe outcome than if the event has a outcome that is less severe.

Now that is not really an appropriate or equitable way to look at an adverse event. All real decisions are made under uncertainty. In making a bet you look at the stakes and the odds, not the outcome, in judging whether it is a good bet. The result, if hindsight

bias is allowed to operate in an unfettered way, is that negligence may be wrongly inferred. That is really scape-goating, an ancient biblical pursuit.

Whenever there is a disaster involving a man-made system, be it healthcare or rail or aviation, our instinctive response is to say:

“who is to blame for this?”

The reason for that is that the alternative is too terrible to contemplate, namely:

“no-one is responsible for it, our system is responsible for it,
our systems are unsafe”

That is extremely disturbing and undermining. So we look for the biblical scape-goat to pile onto his or her back all the faults. In fact, doing this to individuals, which is what scape-goating is, is really a very ineffective way of getting the message out to the other penguins on the ice flow. It just doesn't work in stopping anyone else making the same mistake. Scape-goating implies that errors are deliberate.



[picture]

The person involved in the pictured adverse event had been on back-to-back shifts and on the day this happened the bloke who worked the paint machine was off sick. So he had to do it all himself. How do you think he felt when he saw that ?



How do you think this driver felt when they found about that?

People don't deliberately make errors. Doctors and nurses don't get up in the morning and say “what patient can we kill or injure today?” There is no free will about error – it is not a matter of choice. The systems approach to error recognises that. The systems approach is derived from high reliability organisations in aviation,

nuclear power generation and oil exploration. They have worked this out long ago because their disasters, unlike health care, are large scale and very public. Health care disasters are often on one doctor (or nurse) and one patient.

We have been very slow to learn in medicine but we are learning. The systems approach recognises that the person involved in a medical adverse event is at the apex of a pyramid of system errors that often set them up to make the error with all the certainty of a planned experiment. They are what James Reason (Department of Psychology, University of Manchester UK), who has written a lot about this, calls the resident pathogens in the system. In 5,000 claims against doctors I have seen, time and time again, that there are a minimum of two or three system elements that operate in producing the adverse event.

The problem is when doctors try and look at adverse events in a very critical and scientific way. The medical profession, I think, has been ahead of any other profession in self examination about their outcomes. You only have to go to a mortality and morbidity meeting in a hospital surgical department to see our savage and primeval the criticism. The reason is that there is a culture of perfection - most doctors and nurses want things to go right all the time and that tends to suppress reporting of errors or the admission of them. Of course the culture of perfection is hampered by the fact that imperfection is inevitable.

Retrospect viewers often have problems because of this cultural background, this culture of perfection. There are a number of studies that show that retrospect viewers looking back at a death in hospital consistently over-estimate the probability that that patient could have been saved. They often have excessively optimistic views on how seriously ill patients will go. They are having trouble distinguishing the requirements of the classroom (or the M and M meeting) from the requirements of the courtroom which are:

“were the decisions by the person involved in this event reasonable at the time, under the circumstances as they saw them ?”

The reason why the experts do that is that they are invariably influenced by hindsight bias.

Systems do drift into failure.



The Waterfall train accident that was a good example I think of how systems evolve. When they are set up they, they often start with a series of norms - the culture says this is how things will operate. In the case of trains, for instance, guards are supposed to apply the emergency brake if they see anything adverse in the passage of the train, like excessive speed or something that concerns them. As time goes on of course any guard that pulls the emergency brake and does it at the wrong time, will be criticised by the driver. So the culture drifts to a situation where drivers drive the trains and the guards just blow the whistle and open the doors. That is exactly what happened at the **Waterfall** train accident. It is fine that the safety parameters oscillate so long as they are inside that norm. Trains exceed speed on bends and nothing bad happens. But then you get a combination of the drift of the norm and something bad like a driver dead at the controls with the weight of his leg keeping the dead mans pedal depressed. Then you get a disaster. If you look at that situation, you must not fall into the trap of using the old norm, the original norm of the culture, and failing to consider how far the norm has drifted away from that level.

There is another hazard of hindsight bias. The experts convene and they talk to the people involved about the incident in relation to root cause analysis. Both the experts and the people they are talking to are looking at the incident through the prism of the adverse outcome. So you get people coming out of the woodwork to talk about the personal short comes of the doctor or nurse who was involved. This is exactly what happened in relation to a tragic case a couple of years ago of a young mother who came to Sydney and in being delivered of a baby had an epidural anaesthetic. She unfortunately developed a spinal abscess at the site of the epidural. She went back to Dubbo, diagnosis was delayed and this young mother tragically died. The family's major concern was infection control at the Prince of Wales Hospital. At the coroner's inquiry one of the people they were talking to said "two of the theatre staff said that Dr Collier, (who was the anaesthetist who did the epidural) was known not to wash his hands. Now, that may or may not have been true. Dr Collier vehemently denied it. But even if it was true, what relevance did it have? You would have to establish, first of all, that not washing his hands had some relationship to the development of the abscess. The abscess in fact possibly came from a blood-borne infection from a perinatal wound. Secondly, you would have to do a survey of anaesthetists to know how many did in fact wash their hands, rather than simply donning a pair of sterile gloves, to do what is after all just an injection. I don't know what weight the coroner gave to it and I don't know what the outcome was. But I know it was seriously damaging to Dr Collier and that kind of process, knowing the outcome and then mining facts about someone's personal shortcomings, is a very unreliable process.

Consider for a moment the tort system and medical adverse effects relate to each other. If you define a medical adverse event as an unintended outcome of treatment resolving in prolonged hospital stay, patient injury or death, then it's a very common situation. We know from a number of studies that adverse events both here, the USA and the UK happen in up to 15% of hospital admissions. Many of those adverse events may be minor ones, such as a medication error giving Panadol instead of Aspirin. But nevertheless there is a very big universe of adverse events that injure patients or prolong their hospital stay. The number of those cases that result in litigation is very small:

It is very difficult to imagine that the litigation process can actually have a beneficial effect in reducing adverse events. Plaintiff lawyers will often tell us that it keeps the doctors or nurses careful or honest. Furthermore, if you look at which adverse events actually result in litigation, the most significant selection factor is the severity of the injury - the potential size of the pay-out, not the magnitude of the error. So it is not a very good process for modifying the occurrence of errors.

Error is very ordinary thing. Even the worst disasters, when you look at them; frequently involve just ordinary well-intentioned people making decisions that at the time seem reasonable to them. It does not matter what adverse effect you look at, whether it's in health care or elsewhere, is very often like that. In health care we have a slightly different perspective from the courts in looking back at adverse events. Our over-riding concern is "why did this happen ? how can we stop it happening in the future ?" So the purpose of us looking at adverse events is not to find where people went wrong, it is to find out why their assessment and actions made sense at the time. It is to explain - we are not attempting to excuse what happens and we want to fix the problem not fix the blame.

Medical adverse events do have special requirements. If you are to really understand them and evaluate how they happened, and perhaps who was responsible, you have to consider three unique factors:

- the doctor or nurse concerned (the clinician),
- the nature of the illness
- that the patient has and the constitution of the patient themselves.

You would also need to answer four questions:

- Did the doctor or nurse's actions actually cause the adverse event?
- Was it actually a result of the illness?
- Was it preventable?
- What role did the system play?

That isn't easy in many complicated medical cases. You are looking at a very tangled set of conditions. The clinician, the patient intertwined in among the system, the under-staffing, people being put in positions of responsibility above their level of experience, the lack of equipment. All those systems things also have to be considered. The way that health care is at present trying to grapple with this problem is route-cause analysis.

I'll just spend a few moments talking about route-cause analysis because it is so important in understanding adverse events and in assisting experts to give opinions about causation and so on. It's essentially defined in this way (and this comes from the nuclear power generation industry): the most basic causes of an adverse event that can reasonably be identified and that management has the control to fix. It is a pragmatic definition: "Let's look at what happened here to cause this bad thing that we can actually potentially do something about". The objectives are to make recommendations about those causal factors and to close the loop by doing something to stop them happening in future. The way we do it is to get together a little group of experts as soon as we can after the adverse event; get them to draw up a detailed flow chart; then ask "what, how and why?" but don't focus on "who"; to construct a cause and effect diagram and see where we could put barriers to prevent this error in future.

We usually find at least two or three, sometimes four or five, system elements where we could put a barrier, develop causation statements and recommendations. We try to get the executive of the hospital or the area health services to sign off in a reasonable time and then to go around and check that those barriers have been implemented.

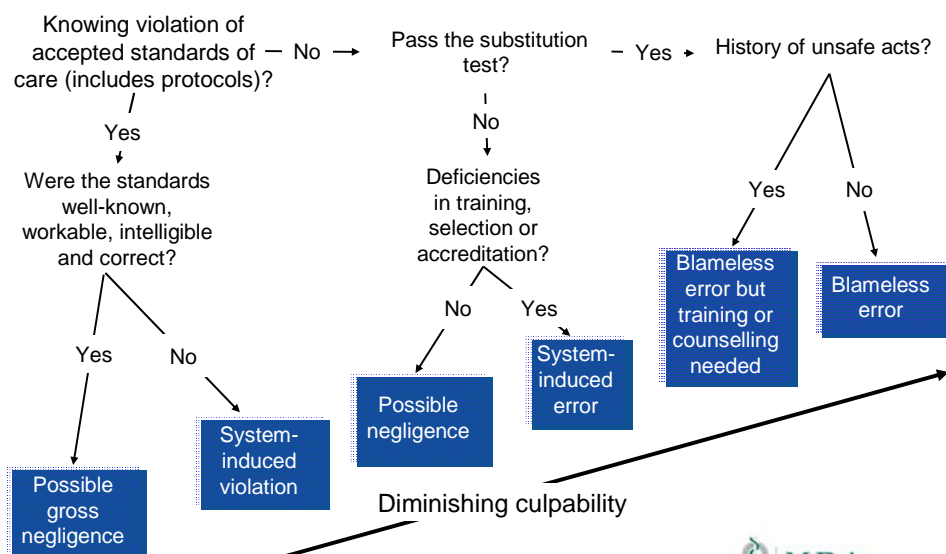
I was involved in giving evidence before the Walker enquiry into Camden and Campbelltown and Bret Walker SC, when we talked about this, quite properly, I think, said:

“well that’s all very well, but I need to know when and how we can identify culpability, because that’s what the newspapers were baying for”.

In the route-cause analysis process we have very clear parameters. Where we would consider route-cause analysis wasn’t appropriate (if there was a criminal act, a deliberately unsafe act, substance abuse was involved or abuse of the patient by the provider) we would say you can assume culpability, it’s not appropriate for us to look at that, that’s a matter a for a court and perhaps medical board proceedings.

What about less obvious culpability where you have got one of these complex intertwined events ? how do we determine that ? One way is to use substitution tests. Ask the individual’s peers: “given the circumstances at the time, can you be sure you wouldn’t have committed a similar unsafe act?” Or the Johnston’s substitution test: you substitute the individual with someone with comparable level of training and experience and say: “do you think it’s likely this new individual would have behaved any differently?” At MDA we also have a rough and ready test that we call a “chap score”: could have happened to any peer; one to five, if you say; well this adverse event, this stuff-up, this could have happened to anybody, we would give that a five; but if it wouldn’t have happened to the rawest registrar or intern, we give it a one. It was a rough way of grading culpability.

Decision tree for determining possible culpability in medical adverse events



Modified from Reason, J. *Managing the risks of organisational accidents*. 1997



You can do it in a more subtle way by this decision tree. You consider: “in the adverse event, was there a knowing violation of accepted standards of care such as protocols ?” If there was, you have to look at “were those standards well promulgated, intelligible, workable and actually right?” It’s surprising how many protocols are absolutely wrong. If they were, then that is possible gross negligence. I do not need to tell this audience that if there is a death that might amount up to a case of manslaughter, but if the protocols did not meet that test then that is a system abuse violation. If there wasn’t a knowing violation of accepted protocol, and this adverse event happened, use the substitution test: “would a peer of similar training and experience behave any differently ? or would you yourself have done this?” If the answer is no, then you don’t stop there and say “well this is an individual who stuffed up” You have to say “why was that individual placed in that position with that level of training and experience, in a position where they would make that wrong decision ? Was there a problem with accreditation? was there a problem with the training program ?”. If no, if the individual was appropriately trained and properly accredited and so on, then that is possible negligence. But if yes, it is the system that has let us down. If the substitution test is passed (that is a comparable individual might well have made the same kind of wrong decision) then you do have to ask “well, does this individual have a history of unsafe acts? if yes, then it is blameless error”. But there may be a problem in the system that required counselling or training, and that is blameless error. So culpability diminishes as you go up that decision tree. That is what we use to determine overall culpability and it’s a guide for us in defensibility or otherwise. We are quite happy to write out a cheque if someone fails this decision tree - to recompense a patient for an injury.

How do we control hindsight bias? One particular issue to do with expert medical evidence often arises where the expert is giving an opinion on, say, some x-rays that were misinterpreted or pathological slides where a serious or perhaps life-threatening condition was missed. It is the problem of conspicuity, that is, how conspicuous the abnormality is in either the x-ray or the microscopic slides.



Here is an x-ray of a hip joint that shows a patient has arthritis, the joint space is thinned and there's this sort of spiky stuff which is an indicator of arthritis in the hip. The history was that the patient fell a few days earlier and had a limp. Now this patient actually has (sub?) fracture of the femur. I don't expect you see it on this projection, but it's an impacted fracture, the force of the fall jammed the bones in together so it's very difficult to see. It is notoriously defective - about 30% of these fractures are missed on the first x-ray. But the allegation was in this case that the radiologist has missed it - on the report, the diagnosis wasn't made. It is possible for patients to walk around on these impacted bones for a while, then it gave way, the patient fell, there was much more damage and the patient had to have a hip joint replacement rather than a simple pinning of the fracture. The issue was: the experts for the plaintiff (three of them) had said "we can see a fracture". That was very difficult for us, we had difficulty seeing it, our experts had difficulty seeing it, and so we tried a little experiment. We showed a meeting of a group of radiologists a whole set of problem x-rays and this one was put in amongst them. They knew that there was a problem with these x-rays (and that is very significant in relation to hindsight bias). We found that six out of 10 could see the fracture but four said there was no fracture visible. There is arthritis in the joint. We then took the x-ray and slipped it into the routine pile of ten radiologists. We were sitting down on a long afternoon in a darkened room in a basement of the hospital looking through x-rays with a request form: fall three days previously, limp. Eight could not see the fracture and the reason was conspicuity - they did not know that this particular x-ray was the subject of litigation and therefore of an adverse event.

Radiologists have said in print, if they are shown an x-ray or sent an x-ray by a lawyer, the first question they ask is: "what was missed on this x-ray?" That is not an objective judgement of whether the original radiologist should or should not have seen it. We have the same problem with microscopic slides. There have been a number of claims made because of failure, for instance, to detect cancerous changes

on a pap smear, for cancer of the cervix. That may be for the patient a fatal error because of the delay in eventual diagnosis of the cervical cancer.

Conspicuousity is a problem that's like this; if you tell someone to look at the night sky and tell you what they see, you know that there is a rather small comet that's visible on this particular night, but you say "look at the night sky and see what you can see". It's quite possible that they will miss an inconspicuous comet. If you say to them "there's a comet", then a greater proportion of people will see it, but if you say "there's a comet just above that mountain peak" then you will have even more people see it, and whether they see it or not will depend on the ambient light, whether there is a moon or whether there is ambient light from the cities nearby.

So, seeing an abnormality in slides that suggests cancer, is a problem of conspicuousity. When you are judging whether a pathologist should or should not have seen it, whether he was departing from accepted standards of care in missing it, you have to consider that particular problem. In fact, the American College of Pathologists are so concerned about this issue that they have suggested this should be the kind of test that should be applied for judging whether an error was made in assessing pap smears. They should be reviewed [blindly?], tucked into the normal kind of reporting without knowledge that there is litigation. Litigation is a clue to an adverse outcome and leads to outcome bias by reasonable scientists. That seems to me a perfectly reasonable suggestion.

To sum up I think there is a need to recognise the effects of hindsight bias on expert evidence. I think there is a case for instructing experts about hindsight bias minimization. But a number of studies in the psychological literature show that you can't get rid of it even when you tell the experts about it and tell them to be careful. It is hardwired into everyone's brains. It is the way we see the world. You can ask them to walk in the shoes or to use substitution tests. If you can, you can conceal the adverse outcome from the experts. It is very difficult, except in the case of objective things like x-rays or slides, to conceal from other experts, the knowledge that there has been an adverse event because they are being asked to review something that has happened in the past. Blind re-screening is important.

So in conclusion, hindsight bias is unavoidable in any retrospective examination. It is a universal **phoneme**. It does distort evaluation of adverse events and sometimes results in harsh judgements that are not justified. It can be reduced by concealing the fact of an adverse outcome from the experts and I think there is a case for including a warning about it in the code of conduct for expert witnesses.