

Making the Wrong Decisions for the Right Reasons

Michael Smithson, The Australian National University

Posted 8/5/2011 on

http://www.bestthinking.com/thinkers/science/social_sciences/psychology/michael-smithson?tab=blog

There seems to be a widespread intuition that if we use a well-reasoned, evidence-based approach to making decisions under uncertainty then we'll make the right decision most of the time. Sure, we'll make some bad calls but the majority of the time we'll get it right. Or will we?

Here's an example from law enforcement. Suppose you're the commanding officer in a local police jurisdiction, and you have to decide how to allocate resources to a missing person case. A worst-case scenario is that the missing person ends up a homicide. Although police are required to treat all missing persons cases seriously, as most do not involve foul play it would be grossly inefficient to treat all missing persons as potential homicides. So, if the missing person isn't found within 24 hours, you'll undertake a risk analysis, considering issues such as whether the circumstances are suspicious or out of character, or there is evidence of the commission of a crime.

What would be your best approach to this risk analysis, and how likely would you be to come to the right decision? A landmark UK study examined 32,705 cases of missing persons in the UK between 2000 and 2002, and determined that 0.6 percent were found dead, although not necessary victims of homicide (Newiss, [2006](#)). This is a very low percentage, and it turns out to be the source of a major headache for you as the commander responsible for deciding what resources to allocate to your case.

You have years of experience, wisdom handed down from seasoned investigators who came before you, and you've read the relevant literature. You know that where a missing person is found to have been a victim of foul play, risk factors include age and sex, involvement in prostitution, last being seen in a public place and an absence of a history of suicide attempts or mental health problems.

So, you're going to make a decision whether to allocate more resources to a missing persons case investigation based on some diagnostic criteria which I'll denote by D. The criteria included in D are indicators that the missing person may have died. There are four commonly used criteria for evaluating how good D is:

- Sensitivity = $P(D \text{ present} | \text{death})$
- Specificity = $P(D \text{ absent} | \text{alive})$
- Positive Predictive Value = $P(\text{death} | D \text{ present})$
- Negative Predictive Value = $P(\text{alive} | D \text{ absent})$

The expressions on the right hand side of these equations are conditional probabilities. For instance, $P(D \text{ present} | \text{death})$ is the probability that D is present given that the person has died. Sensitivity and specificity measure the ability of the model to detect the occurrence or absence (respectively) of deaths. Predictive value, on the other hand, tells us the probability of making a correct diagnosis (death versus no death) based on D.

Now, suppose D has a sensitivity of .99 and specificity of .99 (far better than can be obtained from the otherwise worthwhile predictors identified by Newiss). The next table shows how well D would perform in distinguishing between cases ending in death and cases not involving death.

	D present	D absent		Error-rate		
Dead	196	2	198	0.01	Sensitivity	0.99
Alive	325	32182	32507	0.01	Specificity	0.99
	521	32184	32705			
	pos. pred.	neg. pred.				
	0.3762	0.999938				

Because sensitivity is .99, D misses only $.01 \cdot 198 = 2$ cases involving deaths, and correctly detects the remaining 196. Likewise, because specificity is .99, D absent misses $.01 \cdot 32507 = 325$ cases that do not involve death. That is, there are 325 missing persons with D who will be found to be alive. But 325 is large compared to the number of correctly identified deaths (196). So positive predictive value is poor: $P(\text{death} | D \text{ present}) = 196 / (196 + 325) = .376$. The rate of incorrect positive diagnosis therefore is $1 - .376 = .624$. If you, as commander, decided to allocate more resources to cases where D is present you could expect to be wrong about 62% of the time.

Can these uncertainties be reduced? An obvious and frequently recommended remedy is further investigation into factors that may predict the likelihood of a missing person ending up dead and, conditional on death, being a homicide victim. These investigations could be combined with survival analysis of the kind employed by Newiss, to determine whether there is a relationship between the length of time a person has gone missing and the likelihood that the person ends up dead.

But how effective can we expect these remedies to be? Note that improving sensitivity would have only a negligible effect on positive predictive value. To get to the point where positive predictive value was an even-money bet (.5) would require specificity to be .994. To move positive predictive value to .9 would require specificity to be .9993. Thus the test would have to be incredibly accurate in order to not devote considerable resources to investigations where it was not warranted.

These are unachievable standards. Police will inevitably face a considerable error-rate in making resource allocation decisions regarding missing persons cases. Of course, this does not imply that improving predictions of homicide in missing persons cases is futile, but simply tells us not to expect such improvements to raise the probability of a correct decision to a desirable level.

Mind you, it isn't all gloom and doom. If we consider the false negative problem (e.g., a [Britt Laphorne](#) outcome) it may be possible to obtain a reasonably high predictive value rate without unrealistically accurate predictors. In our unrealistic scenario (with sensitivity and specificity both at .99), negative diagnosticity is .99994. If sensitivity and specificity both were .5 (i.e., coin-toss levels) then negative predictive value would be about $16,253 / 16,352 = .994$. You, as commander, are very unlikely to end up with a Britt Laphorne case which you stand accused of having failed to treat with due diligence. Instead, you are very likely to be chastised by higher-ups and perhaps the media for "wasting" money and resources on cases where the missing person turned up alive and well.

There is an analogous problem in preventative medical testing, where the disorder to be detected occurs at a low rate in the population. For example pregnant women may wish to test for the possibility that their unborn baby has Down's Syndrome. According to an Australian government health assessment document released in [2002](#), when used as a single modality, the standard screening by measurement of Nuchal Translucency in the first trimester has a detection rate for Down's of approximately 73%-82% at a false positive rate of

5%-8%.. Additional ultrasound cues can further increase detection rates for Down syndrome to more than 95%.

The next table shows the most optimistic scenario according to those figures, i.e., sensitivity and specificity of 95%. At the time, about 12.8 per 10,000 births yielded a baby with Downs, so I've included that rate in the table. Downs Syndrome, thankfully, is rare. The result, as you can see, is a positive predictive value of just 2.38%. Given a test result that says the baby has Downs, the probability that it really does have Downs is about 2.4 chances in 100. If these procedures were widely used, there would be many needlessly upset pregnant women—about 97.6% of those whose combined tests came back positive.

	Positive	Negative		Error-rate		
Downs	122	6	128	0.05	Sensitivity	0.95
Normal	4994	94878	99872	0.05	Specificity	0.95
	5116	94884	100000			
	pos. pred.	neg. pred.				
	0.0238	0.9999				

In July last year there was a furore over a [study](#) published in the Journal of the American Medical Association. The study found that of 2176 participants free of HIV infection who received a vaccine product, 908 tested positive even though they had been exposed to the vaccine, not (of course) the virus. That's a false positive rate of about 41.7%. Now, suppose a successful vaccine is developed but it also has this reactivity problem. In any Western country where the rate of HIV infection is low, the combination of a large proportion of the population being vaccinated and tested could be a major disaster. This is not to say that an HIV vaccine would be a bad idea; the point is that it could play havoc with HIV detection.

The chief difference between the medical preventative testing quandary and the police commander's problem is that the negative consequences of the wrong diagnosis fall on the patient instead of the decision maker. Yet this issue is seldom aired in public debates regarding medical testing. Perhaps understandably, the bulk of medical research effort in this domain goes into devising more accurate tests. But hang on—In the Downs test scenario, even with a sensitivity rate of 100% the specificity would have to be 99.87% to raise the positive predictive value to a mere 50%. For a positive predictive value of 90%? Sensitivity would have to be about 99.99%, a crazily impossible target. Realistically, the tests will never be accurate enough to avoid the problem posed by low positive predictive values for rare disorders.

What can a decision maker do? A final point to all this is that in settings where you're doomed to a high decisional error-rate despite using the best available methods, it may be better to direct your energies toward handling the flak instead of persisting in a futile quest for unattainably accurate predictors or diagnostic cues. The chief difficulty may be educating your clientele, constituency, or bosses that it really is possible to be making the best possible decisions and still getting them wrong most of the time.