

# Categorical Nonsense, $p < .05$

Steven Lewis and Denise Kouri

Efficient conversation has no place at lunch. On the surface, the Denise Kouri–Steven Lewis lunch of April 7, 2010, was pure *muda* (wasted time): too many topics, utter disorganization, innumerable tangents. Should we expect the young to memorize the multiplication tables? What percentage of the population has a passable grasp of statistics and calculus? (I didn't say we were fascinating, just inefficient.) This reminded D.K. that her daughter was reading a book on how research and scholarly publication have become tyrannized by a rigid notion of statistical significance – the .05  $p$  value. If a phenomenon has less than one in 20 odds of occurring by chance, it's significant; if one in 19, it isn't. Rigour demands it, case closed.

It popped into S.L.'s head that the same categorical logic governs medicare coverage: services are either in or out, "medically necessary" or dispensable options. When money is tight, look for a province or two to search for healthcare services to delist. Maybe it's tattoo removal or sex change therapy; maybe it's chiropractic treatments or optometric visits. Oregon famously ranked 800 or so services in order of priority and drew a cut-off line where the money ran out. The goal is a clear choice: the service is either in or out. (It must be noted that the experiment, like so many, had mainly middle-class people ranking services to be delivered to the poor. That didn't make it stupid or evil; it just meant that those preparing the Kool-Aid weren't bound to drink it.)

It is important to understand the likelihood of a statistical finding occurring by chance, and likewise laudable to examine the usefulness of healthcare services. The error in both instances is making too much of minor differences and arbitrarily assigning nearly identical phenomena to different categories. Sometimes, as research grant review panels know only too well, arbitrariness is unavoidable. An application rated at 3.93 gets \$400,000, and the next in line, rated 3.91, gets squat. Everyone knows this is conceptually absurd. When the National Institutes of Health in the United States sent out a batch of applications for a second round of review, about 20% of applications changed categories, from funded to unfunded and vice-versa. A score might have changed by a critical tenth of a point if reviewed by another committee, or three hours later in the day, or if one of the primary reviewers had a more soothing voice. But granting agencies exist to pick winners and losers, and there is no fence to sit on. (Well, not the way competitions are currently conducted, but one can conceive of options. A review committee could be given a pot of money and distribute the funds as it sees fit among

all applications rated as worthy of support. Some might get all of the asked-for funding, and others might get less, proportionate to their relative merits.) Sometimes the molehill is a mountain, practically speaking.

Neither statistical practice nor medicare faces such stark imperatives, but both behave as though they do. It is not metaphysical truth but orthodoxy that has made the .05  $p$  value a statistical icon. There is a reason for this, of course – the virtue of clarity has trumped the reality of nuance. However, many findings with a far lower  $p$  value (lower probability of occurring by chance) are utterly meaningless; when you have very large samples with dozens of variables, you can't help but generate statistically significant findings even if you have no plausible hypothesis for why X should be related to Y. Conversely, many findings with a higher  $p$  value (a higher probability of occurring by chance) can be truly groundbreaking. That is why some statisticians argue that the  $p$  values should be reported but not labelled.

The .05 convention is not just a sensible reminder to be cautious about jumping to conclusions. Academic journals often discourage long interpretive sections of papers – the data are supposed to speak for themselves. Some ruthlessly expunge discussion of the potential meaning of a finding outside the conventional significance range (occasionally authors are allowed to set a cut-off point at .1). This runs the risk of overlooking very promising results just over the border from the statistical promised land. But it does more: it narrows the interpretive lens, discourages free-ranging speculation, privileges research designs where large samples are easy to come by and produces far more correlations (X is associated with Y) than causations (X causes Y).

Regardless of where you stand on this debate, the point is that nothing – no Law of Nature, no externally imposed scarcity, no meta-evaluation – makes the convention unavoidable. It is widely adopted because it is easy to apply, transparent and, in a sense, disciplined. But as in all orthodoxies, true believers and others under their sway pay a price.

Likewise, consider how Canadian medicare – and most other health insurance schemes – define healthcare services as categorically worthy or unworthy of coverage. No healthcare services are useful in every circumstance, and few are always useless. Some services of no use to all but a few may be life saving for someone. A sex change operation could literally save the life of a person suicidally unhappy with his or her current identity. The removal of an ill-conceived tattoo from the face of a youth whose appearance would severely restrict employment oppor-

tunities could spell the difference between a productive life and dissolution. It is logically incoherent and inherently unjust to subject citizens to an insurance lottery whereby their coverage status depends not on their level of need and their prospects to benefit from an intervention, but on categorical coverage rules. If Harry needs a cataract operation to improve his vision, he's covered; if Sally needs eyeglasses, she's not. If Milton undergoes predictably useless back surgery, the state pays; yet Mary pays for the back relief she gets from a chiropractor.

The in-or-out decisions end debate precisely where it should begin. The biggest cost problem in healthcare is not an avalanche of voodoo treatments and fraudulent potions. It is the non-essential and sometimes harmful use of services that are effective when used appropriately, or that cost too much while delivering too little. Magnetic resonance imaging (MRI) and computed tomography (CT) scanning are wonderful imaging technologies, but the fivefold increase in their use over the past 15 years has created a culture of overuse that yields about zero at the margins. Good drugs used wisely produce health; the same drugs used unwisely can land you in the hospital.

So why do we play the delisting game over and over again? For one thing, it is easy. If the Canada Health Act doesn't mandate it, we can cut it. Saskatchewan just terminated public cost sharing of chiropractic services, and like many provinces delisted eye examinations provided by optometrists years ago (the poor remain covered). Second, it avoids the touchy subject of appropriateness because that would challenge the sacred ground of clinical autonomy, ask peers to narrow variations in practice, impose accountability where none now exists and demand investment in first-class information technology. Third, it betrays a lack of trust in the ability of providers – be they organizations or individuals – to make sensible and just decisions about the allocation of finite resources.

It is never easy to ration care, and providers may instinctively say no thanks to the more-freedom-and-more-accountability combo. But if the goals are justice and patient-centred care, providers must make context-specific decisions and be prepared to explain why another MRI or a \$50,000 Hail Mary chemotherapy cocktail is neither good medicine nor sound public policy. The alternatives are profligacy and arbitrary third-party decisions – both simple and clear, but neither just.

In-or-out rule making inevitably harms some patients and creates unjustifiable inequities in entitlement. Even worse, it discourages organizations and clinicians from thinking hard about distributive justice and waste, and novel ways to help their patients. Someone has to manage public resources, and budgets must be finite. The question is whether clinicians and healthcare delivery organizations will be active stewards responsible for allocating resources and explaining their decisions. As things now stand, clinicians graze the common on behalf of their patients and wait to see if anyone – peers, administrators – calls them on it. This is a game with a predictable result: major variations in practice, conflict between managers and providers, game playing, grandstanding, the use of money to buy peace and questionable value for money. The squeaky wheels get the grease, and responsible stewards look like chumps.

Nothing good comes from exempting smart people from having to think about and account for their actions, or encouraging them to ignore the realities of limit and hard choices. In-or-out schemes fail the tests of logic and justice. Getting rid of them would signal a commitment to fairness and recognition that arbitrariness gets in the way of good and efficient care. It would broaden system stewardship and bring providers into the deliberative fold. It would require people to think more deeply about value for money, and to be more transparent in and accountable for their decisions. It would give providers the freedom to seek creative solutions to complex patient problems, along with the responsibility to give reasons for what they do. If we really want a more accountable, just and unified healthcare culture, it's a great place to begin.

Madame Minister, tear down this wall.<sup>1</sup>

#### About the Authors

**Steven Lewis** is a Saskatoon-based health policy consultant and part-time academic who thinks the healthcare system needs to get a lot better a lot faster.

**Denise Kouri** is a public policy consultant and program evaluator based in Saskatoon.

1. As unrepentant social democrats, we rarely get to paraphrase Ronald Reagan approvingly.

Inspiring innovation, building relationships and trust,  
and helping to improve shared accountability for a  
more sustainable healthcare system.

[www.kpmg.ca](http://www.kpmg.ca)



AUDIT ■ TAX ■ ADVISORY