

Virtual Mentor

American Medical Association Journal of Ethics
July 2006, Volume 8, Number 7: 469-472.

Medicine and Society

Improving clinical decision making by excising the physician's judgment

by Abraham P. Schwab, PhD

In 1950 a patient with diabetes couldn't test his or her own blood sugar or have it tested by the doctor; a pregnant woman couldn't get an ultrasound picture of her fetus and we didn't know what DNA looked like; the polio and measles vaccines didn't exist and neither did Medicare or Medicaid. Today these are such common aspects of routine medical practice that it's hard to imagine not knowing about the double helix or the perpetual problems of Medicare and Medicaid.

This statement, "In 1950, statistical prediction rules (SPRs) weren't used in medical practice," doesn't evoke surprise at all, but it might if Paul Meehl's work had been taken seriously by the medical community when it was first published in 1954 [1].

What is an SPR?

An SPR in medicine takes a small number of variables specific to a diagnosis or given set of clinical circumstances and uses those variables to produce either a probability for a particular outcome or a recommended course of action. Sometimes SPRs are referred to as "clinical prediction rules" in medical literature, but, as I'll explain later, this can be misleading. One SPR produced for medical practice regards the treatment of prostate cancer. The most effective treatment for prostate cancer (either radiation, resection or a combination of the two) is determined by whether or not the cancer has spread beyond the prostate. A physician typically makes an intuitive judgment based on his or her experience and various clinical and test data such as the stage of the cancer, the prostate-specific antigen (PSA) and the Gleason's score. In 1997, Partin et al. produced an SPR to provide physicians and their patients with a calibrated estimate that would not rely on the intuitive judgment of the physician [2]. Using the table the authors created, a physician can input the clinical stage of the cancer, the PSA and the Gleason's score and quickly determine the probability that the cancer has spread beyond the prostate. Even though this doesn't automatically dictate the course of treatment (physicians will view these probabilities differently for a 70-year-old patient than for a 45-year-old patient), it does provide a sound basis for deciding on the best course of treatment. No longer left to the vagaries of a clinician's intuitive judgment (which may be affected by the fact that the last three patients had prostate cancer that had spread), the physician's recommendation and the patient's decision can be based on more reliable and objective information.

How does one know an SPR is reliable?

In short, one knows by testing it. If patient information plugged into the table mentioned above indicated a 70 percent chance that the prostate cancer had spread, that prediction would be verified using new cases. In their study, Partin et al. used 4,133 cases from three different centers to produce their SPR. Evidence of this SPR's reliability was then produced through a "bootstrap estimate." Partin et al. drew random samples from the original 4,133 cases and tested the SPRs prediction against the known clinical outcome in each case. More evidence can be gathered by systematically evaluating how well the SPR predicts the spread of prostate cancer in new cases; that is, cases which weren't used to produce the rule. For example, if the researchers had produced the SPR using data from only 2,500 cases and then tested it on the approximately 1,500 remaining cases, they would have had even more evidence that the rule worked.

Why would physicians want to use an SPR?

Before physicians and other experts make recommendations, they gather information and interpret it. When an oncologist suggests resection of a patient's tumor, that recommendation is based on raw data—perhaps from an MRI, a biopsy or a mammography—and an interpretation of that data. Looking over these reports, the oncologist and surgeon must decide whether to recommend a lumpectomy or mastectomy or other treatment.

Presumably, the advantage of expert judgment in such decisions is that the expert has "been here before." It's a comfort to rely on the expert's experience—she's interpreted this kind of x-ray before, probably remembers how it turned out in the past and thus seems to be in the best position to make the most accurate interpretation.

That experts make better judgments than novices is not generally challenged, and I will not challenge it here. But studies have shown that decision models based on an expert's past decisions outperform the expert himself when applied to new decisions [3]. Presumably this is because even the best expert, like the jump shooter in basketball who sometimes lets her elbow drift away from her body, is inconsistent from time to time. Most importantly, many other studies have shown that SPRs generally match or outperform the decisions of the best experts [4]. The great boon of SPRs and the optimism about their potential benefit for medical practice is that every patient can have his or her treatment guided by decisions that match or improve on the intuitive judgments of the top experts.

SPRs and computer-assisted decisions

Computer assistance is available for some of the complicated calculations that are often part of clinical decision making. For example, a physician who wants to know the odds that a patient's positive test results indicate actual disease can refer to free online diagnostic test calculators [5]. An example used by Gigerenzer and Hoffrage can offer helpful explanation here. Let's say that 1 percent of 40-year-old women have breast cancer and that mammography has 80 percent sensitivity (gives a

positive result to 80 percent of the women with breast cancer) and 90.4 percent specificity (gives a false positive to 9.6 percent of the women who don't have breast cancer). Now, if a 40-year-old woman gets a positive test result, what's the chance she has breast cancer? [6]

A physician who knows the Bayesian algorithm for posterior probability can do the math: 1 percent multiplied by 80 percent divided by (1 percent multiplied by 80 percent plus 99 percent multiplied by 9.6 percent). Alternatively that physician could rely on a program like the diagnostic calculator mentioned above to compute the answer. Making use of such a program is referred to as computer-assisted decision making [7].

Computer-assisted decision making is not the same thing as following an SPR. The difference is that a computer-assisted decision is one for which a computer has executed a complicated calculation; an SPR, on the other hand, is a heuristic or easily remembered rule that the physician, once she has a few other pieces of information, can quickly translate into a prediction or recommendation. It's true that the SPR for prostate cancer gives a percentage chance that the disease has spread and the program described above gives a percentage chance of disease presence following a positive mammogram; the distinction between the two lies in the root of the prediction. The computer-assisted decision takes a test result and calculates or determines its meaning. The SPR takes several pieces of information (clinical stage, PSA and Gleason's score) and predicts the possibility of specific clinical results.

The line between computer-assisted decisions and SPRs won't always be bright, just as it isn't in the above examples. Indeed, in some cases, a physician might use computer assistance to get one piece of information (e.g., the Gleason's score) and then plug that information into an SPR. A clearer example of an SPR without computer assistance is the Ottawa Ankle Rule. This is a simple rule that can tell a podiatrist or other physician whether or not to get an ankle or foot x-ray following a blunt trauma to the ankle. When using this SPR, five pieces of information about a patient's foot and ankle tell a physician whether or not a foot or ankle x-ray is needed [8].

As I mentioned earlier, SPRs have also been described as clinical prediction rules in the medical literature. This can be misleading because some clinical prediction rules are straightforward SPRs (e.g., the Ottawa Ankle Rule), while others are computer-assisted decisions. It would be a mistake, then, to assume that all clinical prediction rules are SPRs.

The future of SPRs

Despite the conclusions of the studies, few SPRs have been used in medical practice from Meehl's day until the present. I have noted above one for prostate cancer. A number of others can be found on the Mount Sinai School of Medicine Web site [8]. Given the fifty years since Meehl's work, you might think more would be in use, and I think more should be.

Without an effort to produce and use SPRs in clinical care, physicians restrict the future of the art of healing by subjecting powerful evidence-based therapies and diagnostics to the inconsistent intuitive judgments of its practitioners. Dawes has noted that “the ineffable, intuitive clinical judgment is very difficult to challenge—at least, not without an extensive statistical study to assess its bias” [9]. We can only hope that with robust research conclusions illustrating the predictive reliability of SPRs we will overcome our blind faith in intuitions.

Notes and References

1. Meehl PE. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, Minn: University of Minnesota Press; 1954.
2. Partin AW, Kattan MW, Subong ENP, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA*. 1997;277:1445-1451.
3. Goldberg LR. Man versus model of man: a rationale, plus some evidence, for a method of improving on clinical inferences. *Psych Bull*. 1970;73:422-432.
4. In his 2005 article, Dawes cites a number of the 135 existing studies showing that SPRs are as good as or better than the best experts. Dawes RM. The ethical implications of Paul Meehl’s work on comparing clinical versus actuarial prediction methods. *J Clin Psychol*. 2005;61:1245-1255. Note: at times in this article a copyeditor mistakenly replaced “clinical versus actuarial” with “statistical versus actuarial.”
5. Schwartz A. *Diagnostic Test Calculator*. Available at: <http://araw.mede.uic.edu/cgi-bin/testcalc.pl>. Accessed June 23, 2006.
6. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. *Psychol Rev*. 1995;102:684-704.
7. Either way, this hypothetical 40-year-old woman with a positive mammography has an 8 percent chance of actually having breast cancer.
8. Mount Sinai School of Medicine. *Ottawa Ankle Rule: X-ray After Blunt Trauma?* Available at: <http://www.mssm.edu/medicine/general-medicine/ebm/CPR/ottawaFrame.html>. Accessed June 23, 2006.
9. Dawes RM. *House of Cards: Psychology and Psychotherapy Built on Myth*. New York, NY: Free Press; 1994:104.

Abraham P. Schwab, PhD, is the senior fellow in the Institute for Ethics at the American Medical Association in Chicago, Ill.

The viewpoints expressed on this site are those of the authors and do not necessarily reflect the views and policies of the AMA.

Copyright 2006 American Medical Association. All rights reserved.