Imaging, Screening, Artificial Intelligence, and the Diagnostic Dilemma: An Epidemiologist's Response

John C. Lane

Follow this and additional works at: https://scholarlycommons.henryford.com/hfhmedjournal

Part of the Life Sciences Commons, Medical Specialties Commons, and the Public Health Commons

Recommended Citation
Available at: https://scholarlycommons.henryford.com/hfhmedjournal/vol34/iss2/20

This Article is brought to you for free and open access by Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Henry Ford Hospital Medical Journal by an authorized editor of Henry Ford Health System Scholarly Commons.
Imaging, Screening, Artificial Intelligence, and the Diagnostic Dilemma: An Epidemiologist’s Response

John C. Lane, MD, PhD*

Issues in imaging, screening, and information processing discussed in a special issue of this Journal (Henry Ford Hosp Med J 1985;33:65-148) have implications for the decision-making algorithms of all clinicians who use those imaging and screening techniques. Epidemiologic and psychological research show that clinicians, like other professionals, do not obey the laws of conditional probability in their judgments of risk or outcomes under uncertainty. Although physicians cannot be expected to make complex probabilistic calculations every time they receive the result of a screening test, teachers of medicine should present algorithms that would allow the physician-in-training to make efficient use of information from all sources, including screening tests. In the long run, more formal training of both clinicians and their teachers in basic epidemiology and biostatistics, especially Bayesian probability, might yield more efficient use of information from the complex screening systems now available. (Henry Ford Hosp Med J 1986;34:144-6)

A special Journal issue on medical imaging (Henry Ford Hosp Med J 1985;33:65-148) addressed relatively recent strategies in screening and medical information processing which have major importance for the clinician as well as for the radiologist. Ackerman and Burke’s article on artificial intelligence (1) and Keller and Watt’s article on mammography as the standard for imaging and screening for breast carcinoma (2) both address a more fundamental, underlying problem that all physicians face. The pediatrics faculty at Henry Ford Hospital taught this physician, in the course of his clinical training, that the license to practice medicine is a license to make decisions with less than complete information. How the physician handles the incomplete information available to him, then, may be crucial in effective diagnosis and treatment.

Thus, the basic issue of judgment under uncertainty is possibly the fundamental problem of medicine. This applies to the surgeon, internist, gynecologist, psychiatrist, and pathologist as well as to the radiologist, pediatrician, and epidemiologist. This paper examines this basic issue of judgment under uncertainty, with references to literature in epidemiology and the social sciences and to some of the older medical literature, which illustrate some of the difficulties that must be faced both in finding algorithmic solutions to the diagnostic dilemma and in other applications of medical information processing.

The job of intelligent problem solving in medicine is not mere listing, sorting, and moving strings of information. Judgment under uncertainty and the efficient use of decision-making algorithms depend, as Ackerman and Burke (1) noted, on the rules of conditional probability, including Bayes’ theorem. With the exception of a single lecture in a freshman course in medical genetics, this author was never taught Bayesian probability in medical school. Bayes’ theorem, however, did receive a scant net of attention in the course’s text (3). It is not surprising that Eddy (4) found physicians to be unaccomplished in using information from screening tests in making diagnoses.

The example that Eddy used to illustrate clinicians’ difficulties in using screening tests to make clinical decisions is the same example cited by Keller and Watt (2) as the standard for screening and imaging; the use of mammography in screening for breast carcinoma. Eddy (4) developed substantial evidence which shows that physicians tend to erroneously but intuitively assume that the probability of cancer in a patient with a positive mammogram [P(pos|ca)] is approximately equal to the probability of a positive mammogram in a patient with cancer [P(pos|ca)]. This will often contradict Bayes’ law, which states that:

$$P(ca|pos) = \frac{P(pos|ca) P(ca)}{P(pos|ca) P(ca) + P(pos|no-ca) P(no-ca)}$$

As apparent from this equation, the laws of probabilistic decision-making are counterintuitive. Intuitive judgments and decisions made on the basis of those judgments tend to be badly biased. This is a particular problem when physicians have to use base rates in relevant sections of the population as the values for the prior probabilities required in Bayesian analysis. Most physicians do not have a clearly formed idea, either from the epidemiologic literature or from their years of experience of seeing patients with similar risk factors, what P(ca) and P(pos) might be in the aforementioned equation.

Kahnemann and Tversky (5,6) have shown that laypersons tend to ignore base rates and sample sizes in intuitively assessing risks or probabilities. Later research showed that physicians, educational psychologists with advanced statistical training, business school graduates with training in decision theory, and others are also prone to judgment using their decision making.

A diagnostic algorithm, by which it generates these decisions, is caused so much by a mammogram as to produce the predictive mammogram and surgical necessity firmification of ‘‘medical lit­ ings’’ (8). It is not a sure negative n which is most likely to be true in a (2), or more (9), or in a way.

The use of breast malignancy, breast family history, and early Eddy (4) claimed that P(pos|ca) was 8%. Usual-Ability of the patient is about one of the critical factors. If the base is b, the patient b then would be true in a mammography.

More specifically (12), P as high as 8 years of s brion D mammography would have been mammmogram exposures. With the view of conditions D rate of breast st. would more with uncorre-
and others all tend to depend on the same intuitive rules for judgment under uncertainty which laypersons use in making their decisions.

A diagnostic or screening procedure is useful only if the clinician ordering the procedure knows how to use the information which it generates. Perhaps it is the lack of training in information processing and probabilistic decision-making that has caused so many physicians to ignore the information provided by a mammogram. Eddy (4) cited clinicians' confusion between the predictive accuracy (or positive predictive value) of the mammogram \( P(\text{caipos}) \) and the retrospective accuracy of the mammogram \( P(\text{pos|ca}) \) as the likely reason for the medical and surgical literature's insistence that "biopsy is as much a necessity for confirmation of X-ray findings as it is for the confirmation of physical signs" (7). Similar statements from older medical literature stated that "any palpable lesion requires verification by excision and biopsy regardless of the X-ray findings" (8), and that "while mammography is usually definitive, it is not a substitute for biopsy" (9) [cf (10)]. If this is the case, a negative mammogram could never prevent a biopsy of even the most obviously benign lesion on the mammogram. This would be true in spite of the risks and drawbacks which Keller and Watt (2) cited, including potential disfigurement, the risk of infection, and the costs of the procedure.

Using a classical X-ray mammogram and a patient with a breast mass who has about a 1% chance of having a malignant lesion, based on a clinician's experience with women of similar family histories, ages, medical histories, and physical findings, Eddy (4) used statistics from 1966 (11) to show that though \( P(\text{pos|ca}) \) is as high as 0.792, \( P(\text{caipos}) \) is only approximately 8%. Using the same statistics, Eddy also showed that the probability of cancer and a negative test occurring in the same patient is about 200 times less than the probability of a positive test in a patient without cancer.

If the prior probability of breast cancer in a woman with a palpable breast mass is as high as 8%, a positive mammogram would raise the \( a \text{ posteriori} P(\text{caipos}) \) to 40%, while \( P(\text{caineg}) \) would still be slightly less than 1%. Under these conditions, mammography could be efficiently used to guide decisions as to whether to biopsy in some patients.

More modern data on the validity of mammography would further strengthen this argument. In the series by Mygind et al (12), \( P(\text{caipos}) \) ranged from 63.2% to 68.2%, and \( P(\text{pos|ca}) \) was as high as 91.5%. \( P(\text{caineg}) \) was only 1.7%, and \( P(\text{neg|ca}) \) was only 8.5%. Clearly, mammography has improved in the 20 years since those studies on which Eddy had based his conclusions. Dodd (13) has also documented the improvements in mammography which have resulted in better contrast, shorter exposure time, lower radiation dose, and automated processing. With these improvements, the true positive rate in Dodd's review of the recent literature and the Breast Cancer Demonstration Detection Projects was 87% to 89%, with a false-negative rate of 8.6% to 12.8%. Obviously, if Eddy had used these modern statistics, his calculated predictive value of mammography would have been higher, and the mammograms could be used more frequently to guide decisions as whether or not to biopsy.

With the currently available techniques of ultrasound, thermography, diaphanography, computed tomography, and digital subtraction angiography (2), the predictive value of imaging techniques should increase still further.

Sufficiently well-developed algorithmic reasoning, where the decision trees carefully illustrate every potential branch point, may be used to guide judgment under uncertainty. These strategies have been used in the past in training physicians' assistants, nurse clinicians, and military medical personnel (14). Use of these algorithms by physicians in clinical medicine has brought the research diagnostic criteria for mental disorders (15) into their maturity in the Diagnostic and Statistical Manual of Mental Disorders (16).

Because Bayes' theorem is inadequate when reasoning from incomplete knowledge (about prior probabilities, for example) and because the complexity of probabilistic calculations increases as multiple hypotheses are tested, it is certainly not reasonable to expect that clinicians will replace their intuitive clinical judgment with formal probabilistic calculations whenever they use information from screening tests. Physicians' handling of that information efficiently will require systematic protocols of clinical reasoning in which no branch or algorithm in diagnosis or treatment is neglected as a result of lapses in clinicians' intuition. When physicians and their coworkers become fluent with those protocols, they can then undertake more formal statistical and epidemiologic reasoning to develop improved strategies for judgment under uncertainty.

While it is optimistic to hope that clinicians will be trained to mentally perform complex probabilistic calculations as they make use of a diagnostic test, it is not unreasonable to expect that instructors of medicine and surgery will teach algorithms for the use of screening tests, imaging techniques, and decision strategies that make the most efficient use possible of all available information. Teaching those algorithms may actually develop "clinical reflexes" and habits that guide physicians-in-training to make efficient judgments under uncertainty.

If the algorithmic strategies used by clinicians for making decisions under uncertainty need refinement and more use of efficient information-processing strategies, it would not be an optimal strategy to develop artificial intelligence devices (hardware or software) which were simply designed to imitate the current habits of an intelligent clinician. The basis for the decision-making strategies of the artificial intelligence devices will need to be, to the extent possible, free of the biases of intuitive rules of thumb for risk assessment and decision-making that hinder physicians in effectively using information from complex screening procedures.

Such artificial intelligence devices and programs must therefore improve upon and not merely imitate the clinically trained clinician's use of probabilistic reasoning. Turing's (17) criterion of the machine's ability to imitate intelligent communication indistinguishable from the clinician's communication will not be successful for the purpose. The machine algorithms will need to set examples for the intelligent clinician and teach the problem-solving approach of probabilistic judgment under uncertainty in which clinicians have been less than optimally trained.

Machines programmable both to learn and to teach are already in development for other applications. Ackerman and Burke (1) discussed the use of such machines in quantitatively and probabilistically arriving at a radiographic diagnosis when
used with some complex imaging techniques. Given this background, Turing's criterion of being able to imitate intelligent (in this case, clinical) reasoning will probably not be adequate for the future of medicine. In the short run, simple teaching algorithms which train clinicians in the use of all available information will be an appropriate goal in solving the diagnostic dilemma.

In the long run, as the technology of medicine becomes ever more complex, more formal training of clinicians and their teachers in probabilistic decision-making and Bayesian and classical statistics may be required. Such epidemiologic concepts as what a false-positive test and a false-negative test mean will be important, as will the concepts of a test's sensitivity, specificity, and positive predictive value. These concepts will need to be understood by all physicians, not just epidemiologists, and they will need to be used in the processing of clinical information and in decision-making. Clinicians and their teachers will need to know base rates (or have them available), know how to interpret 2 x 2 tables, and know something about risks in populations to function effectively.

Acknowledgment

The author is supported by training grant #5 D33 PE16000-03 from the US Department of Health and Human Services, Public Health Service, Division of Medicine, Tulane University School of Public Health & Tropical Medicine.

References