April 2003

Clinical decision making Part I: errors of commission and omission

F White
Aga Khan University

D Nanan
Aga Khan University

Clinical Decision Making Part I: Errors of Commission and Omission

F. White,D. Nanan (Department of Community Health Sciences, The Aga Khan University, Karachi.)

Introduction

Clinical decision making refers to any act of diagnosis that leads to a decision regarding prognosis, treatment, referral, or counseling. Diagnosis itself passes through a series of stages, beginning with the assessment of symptoms. However, the meaning of symptoms varies greatly with the context. For example, diagnostic outcomes for cough presenting in general practice differ from those seen in the context of a chest clinic. Persons presenting with severe headache at a neurology clinic are more likely to have a brain tumor than those presenting with a similar complaint in an emergency clinic.

In light of presenting symptoms, the clinician then proceeds to the stage of diagnostic hypotheses, generally restricting the serious possibilities to less than four (said to relate more to the limitations of short term memory than to the intrinsic validity of any such restriction).1 He or she then attempts to use available evidence to differentiate among these hypotheses. Ideally, this leads to isolating a single diagnostic entity. Consequent upon this, a prognosis is made, and a course of treatment identified. However, potential for error exists at all stages in this process, and actions, whether valid or invalid, have consequences.

The Use and Abuse of Frequency Statistics

That statistics influence perceptions in medical practice has long been recognized. Almost 50 years ago, one exponent told the following story2 of a friend who came one day and asked, "A member of my family has to be hospitalized. How can I tell what hospital to put him in?" Being head of a hospital activity study at the time, he answered, "We'll just see which hospital is the safest to go into. We have 15 hospitals which have been contributing some data. Let's see which reports the greatest percentages of recoveries on discharge." He drew attention to Hospital #1, where 92% of patients recovered and Hospital #14, where only 28% recovered, and stated: "Obviously, there isn't any sense in going to Hospital 14, you've only got a 1 in 4 chance of getting out of there cured!"

The author then conceded the inherent oversimplification, pointing out that if frequency statistics are to help in decision making, words must mean the same in every setting, and statistics must be accurate and reflect actual practice.2 Since then, even more important issues have been recognized, such as the critical relevance of caseload characteristics: is it a general hospital or a referral hospital? is it private or public sector? do patients arrive with similar or different prognostic profiles? These considerations pertain to what may be termed the "prior probability" of differing outcomes.

For example, in most countries public sector health care administrators are required to balance demand for different types of services against the available and finite supply of resources. This usually results in a limited availability and turnaround for certain procedures and services while maintaining affordability for a wide cross-section of the catchment population. Therefore, is it ever reasonable to compare health outcomes of a public sector hospital with those of a private sector hospital, without first taking account of pre-admission prognosis? The issues include: socioeconomic spectrum of people accessing health care from a given source, the stage at which disease presents (is it more or less advanced), and how much will cost of medication be a barrier. These same issues are equally relevant when comparing one physician's practice with another; those serving wealthier patients will inevitably achieve apparently more favorable outcomes than those who are dealing mostly with poor patients. Frequency statistics therefore are merely numbers: for interpretation they must be placed in context, and only after controlling for the relevant factors can valid comparisons be made.

The Role of Moral Judgements in Medical Diagnosis

The human factor in clinical decision making in the face of uncertainty has also received attention. Especially in the absence of evidence, informal norms (or "rules of thumb") are often adopted in medical diagnosis. The most important norm has long been recognized: "that judging a sick person well is more to be avoided than judging a well person sick".3 By contrast, an opposite analogy of common law in the face of uncertainty holds that: "A man is innocent until proven guilty"; in effect, the judge must find compelling evidence of guilt beyond reasonable doubt. The error to avoid is to erroneously convict, as in the dictum, "Better a thousand guilty men go free, than one innocent man be convicted". Unlike common law, whose foundations rest on a null hypothesis of innocence, the medical profession traditionally abides by a converse rule. Because medicine is concerned mostly with people in the "sick role", the null hypothesis, Ho ("H nought"), is usually taken as no difference between the individual being assessed and others with the disease; after all, physicians deal most of the time with people who are not well, such that "not well" is the operational norm. Little wonder therefore that the medical model is more noted for its obsession with illness than its interest in health; it is virtually designed to do so! However, although the magnitude of the bias towards intervention in a given case may be small, if multiplied across a population it can have effects of large magnitude. Interestingly, compared with law, medicine offers more leeway with its rule, such as "When in doubt, observe and/or delay your decision". This option however, is acceptable only when delay is unlikely to lead to catastrophic outcome, and may entail an offsetting consideration such as potential adverse consequences of taking immediate action.

The Role of Statistical Inference

To examine this question more closely, we invoke inferential statistics. Taking a clinical null hypothesis (i.e. no difference between the person being assessed and others with the disease), and basing our decisions on clinical/diagnostic testing, we are liable to make two types of errors. Based on their test results, we may decide that a person does not have the disease, and thus reject the clinical Ho, when in fact Ho is true (i.e. the individual is truly sick); this is construed as a Type 1 error. Alternatively, deciding that the person has the disease thereby accepting Ho, when the individual is truly well is termed a Type 2 error. These two errors can occur by chance alone, and with predictable frequency depending on the "normal range" we set for the test results for the population under consideration. To repeat, to reject a null hypothesis when the null hypothesis is true is called a Type 1 error. By contrast, to fail to reject, or to accept the null hypothesis when it is false is termed a Type 2 error. Such errors occur routinely in clinical practice. The problem is that it is impossible to tell if we are making a Type 1 or a Type 2 error.

From a purely neutral perspective, Ho may be taken on the basis of either health or disease, but such neutral positions rarely apply in health care. In fact, most physicians learn early in training that it is far more culpable to dismiss a truly sick patient as being well, than to classify a well person as sick. Theoretically, what one takes as Ho is assumed at the outset: Ho may be taken to mean that disease is present unless demonstrated otherwise, or conversely, Ho may be taken to mean that disease is absent. More often than not, in clinical medicine (for the reason of medical morality as just alluded), Ho assumes that disease is present. A common exception to this is screening, where falsely "labeling" someone with a disease is now recognized as an issue; in this situation, the null hypothesis of "health" is usually taken, and the implications of Type 1 and Type 2 error reverse.

Some Classics from the Literature on Diagnostic Errors

There are classics in the medical literature which illustrate the enormous force of the decision rule when based on the null hypothesis of disease, and the greater acceptability and predominance of Type 2 errors in clinical decision making. For example, as early as 1959, Garland summarized findings from a study of 14,867 radiographic films for tuberculosis.4 In this series there were 1,216 positive readings which turned out to be clinically negative (Type 2 errors) and only 24 negative readings which turned out to be clinically active (Type 1 errors). Type 2 errors in this series therefore were found to be about 50 times more acceptable than that of Type 1 errors!

Rules of thumb however, can propagate Type 2 errors to virtually outrageous proportions. To illustrate, take Bakwin's classical 1945 study of tonsillectomy for 1,000 school children.5 Of these children, 611 had had their tonsils removed. The remaining 389 were then examined by other physicians, and 174 (44.7%) selected for tonsillectomy. This left 215 children with apparently normal tonsils. Another group of doctors was then assigned to examine these 215 children, and 99 of them (46%) were judged in need of tonsillectomy. Still another group of doctors was then employed to examine the remaining children, and nearly one half were recommended for operation.

These examples illustrate several issues: 1) the null hypothesis of disease and therefore Type 2 error dominates; 2) clinical diagnostic practices can amount to little more than ritual (in the previous example, an apparent "rule of halves"); 3) errors propagate; 4) it takes courage to buck a trend in medical practice; and 5) the surgical equivalent of the maxim "when in doubt, diagnose illness", could be stated as "when in doubt, take it out". Or perhaps more generally, "it is better to be damned for doing something than for doing nothing". This is not medical science.

The history lessons do not end there. For many years after this study was reported, tonsillectomy remained popular. However, this was also accompanied by great ignorance on the physiological role of the tonsils, assumed by many to be residual lymphoid tissue that performed little useful function. Only after large scale epidemiological studies of persons whose tonsils had been removed, did it become apparent that such individuals had a statistically higher risk of childhood leukemia, Hodgkins disease, inflammatory bowel disease and rheumatoid arthritis, all suggesting a relationship to the importance of a fully intact immune system. While more work is required to elucidate such associations, they caution us to avoid arbitrary rules of thumb. Although there are legitimate reasons for tonsillectomy, which is done less frequently today, a recent study of variations in surgical rates revealed a higher risk of undergoing tonsillectomy among the general population when there were no medical friends or relatives than when there were.6 At least some of the general public, at least those who are friends and relatives of doctors it seems, have got the message.

However, aside from some surgical procedures, it might be argued that medical diagnosis, unlike law, is not an irreversible act that can do untold damage to the status and reputation of the patient. Yet, is this assumption warranted? Consider the physician who suspects epilepsy in a truck driver: his patient will probably never drive a truck again if the diagnosis is made known to the licensing authorities. On the other hand, if he does drive, and has a seizure on the road, many people could be hurt. Similarly, some psychiatric diagnoses can have long term effects on a person's social standing. Consider also the dentist who tests positive on screening for hepatitis B or HIV. Such errors can indeed have consequences, and for this reason confirmatory tests are required.

In fact, it may improve our clinical rigour if, in addition to clinically interesting cases at Grand Rounds, we also had regular sessions focusing on errors in clinical practice. These would consider conditions under which errors are more likely to occur, what type of errors, and their consequences, taking into account the condition being assessed, and type of physician involved, including type of training as error distribution will differ across disciplines, and none will be found immune to errors. Similarly, the individual circumstances of patients, such as educational background and ability to pay, will have a bearing on error rates, as these may affect their pretest likelihood of disease and quality of care. The organizational setting is also relevant, as patient volume, financing system, and cultural setting are also likely to influence outcome.

In addition, system differences can profoundly affect results. For example, the US system of "fee for service" and high litigation rates produces more "errors of commission" related to excessive numbers of procedures, whereas in the UK system of "capitation", where physicians are paid salaries in proportion to the number of registered patients (regardless of clinical activity), there are more "errors of omission", or failure to take action. In either setting this can be viewed as "rational economic behavior".

Acknowledgements

This review and formulation of the issues introduced in Part I and the accompanying

exercise on the effect of varying disease prevalence on predictive value of positive test results detailed in Part II, were developed for the Post-Graduate Medical Education Conference, Complexity Science & Health Care, May 31 and June 1, 2002, The Aga Khan University, Karachi.

Refrences

1. Tomasi P. Logic and diagnostic. Available at: http://www.ul.ie/~philos/vol3/gnostic.html. Accessed May 10, 2002.
2. Slee VN. Statistics influence medical practice. Bull AmColl Surg 1954;39: 394-402.
3. Scheff TJ. Decision rules, types of error, and their consequences in medical diagnosis. Behav Sciences 1963; 8: 97-107.
4. Garland LH. Studies of the accuracy of diagnostic procedures. Am J Roentgenol Radium Ther Nuclear Med 1959;82:25-38.
5. Bakwin H. Pseudocia pediatrica. N Engl J Med 1945;232:691-7.
6. Ruiz I, Hernandez-Aguardo I, Garrido P. Variation in surgical rates: a population study. Med Care 1998;36:1315-23.