CLINICAL EPIDEMIOLOGY ROUNDS

Clinical disagreement: II. How to avoid it and how to learn from one's mistakes

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS,
McMASTER UNIVERSITY, HAMILTON, ONT.

In the first of our clinical epidemiology rounds (*Can Med Assoc J* 123: 499, 1980) we presented several cases of clinical disagreement over patients' histories, physical examinations, diagnostic test results, and diagnoses and therapeutic recommendations, and then described 10 reasons why clinical disagreements occur. This round will conclude the consideration of clinical disagreement by describing six strategies to prevent its occurrence and by presenting a plan for continuing self-education in the improvement of diagnostic accuracy.

When does clinical disagreement really matter?

The strategies for preventing or reducing clinical disagreement demand the time and effort of at least one and sometimes several clinicians. They should be reserved, therefore, for the evaluation of those crucial items in the history, physical examination and diagnostic evaluation that determine a patient's diagnosis, prognosis or management.

Case 1

A 50-year-old man is to undergo elective cholecystectomy and on hospital admission he states that he thinks he "may have had a heart attack a few months ago".

Comment: Both the fact and timing of this cardiac event are crucial. If he did have a myocardial infarction and it occurred within the previous 3 months his risk of having another one postoperatively approaches 30%,[1] whereas postponing his elective operation until 6 months after his infarct reduces this risk to about 4%.

Case 2

A child with scarlatina, fever and arthralgia has a questionable new heart murmur.

Comment: The cardiac examination is crucial in this case. If this child does have carditis the Jones criteria[2] for acute rheumatic fever are met, with all their implications for recurrence, prognosis and daily antistreptococcal prophylaxis.

Case 3

A young athlete develops calf-swelling following immobilization for torn knee ligaments and a venogram reveals "equivocal" evidence of deep vein thrombosis in the ipsilateral thigh.

Comment: In this case, the interpretation of the venogram is crucial. If this patient has proximal deep vein thrombosis the season is over and a 7- to 10-day course of full-dose, intravenous heparin plus 3 months of oral anticoagulants are required.[3]

In these cases a specific item in the history, in the physical examination or in the interpretation of a diagnostic test has a crucial effect on the patient's prognosis and treatment. When such crucial elements arise in the clinical evaluation of a patient, steps must be taken to prevent or at least minimize both inaccuracy and inconsistency in determining their presence and significance.

The prevention of clinical disagreement

Strategies for preventing or minimizing clinical disagreement are summarized in Table I; they have been taken from several sources, including an important series by Feinstein.[4-7]

Match the diagnostic environment to the diagnostic task

Select a site with the appropriate light, heat, silence and privacy for the diagnostic tasks to be carried out. This may mean moving the patient to a more suitable place; if such transport becomes a routine prerequisite for a proper clinical examination, modification of the ex-

Reprint requests to: Dr. R. Brian Haynes, McMaster University Health Sciences Centre, 1200 Main St. W, Hamilton, Ont. L8N 3Z5

amining room is in order.*

Seek corroboration of key findings

This can be accomplished in four ways.

Repeat key elements of your examination: The justification for repeating your history and physical goes well beyond the hope that "something may turn up". First, your patient's memory, jogged by the earlier interview, may now have recalled and reorganized important historical events or key precipitators and alleviators of cardinal symptoms. Indeed, this phenomenon may partially explain why the "attending" physician's history is often more informative than the initial history elicited by the clinical clerk.

Second, biologic variation and, especially, regression toward the mean (the tendency for extreme laboratory results or physical findings to revert toward less extreme results or findings on repeated examination[9]) may have occurred in the body systems under scrutiny, permitting a more clear-cut decision on whether prior findings (such as blood pressure) were or were not normal.

Third, if you carried out the previous examination when you were fatigued, a second examination when you are rested may uncover key findings previously missed.

Finally, a repeat examination may disclose key items that simply were overlooked on an earlier exam, a phenomenon that is well documented in reading chest x-rays.[10]

Corroborate important findings with documents and witnesses: Key items in the history and physical are often documented in prior clinical or health records. Did your patient have neonatal asphyxia? Was she diabetic during her first pregnancy? Did he receive anticoagulants when taken sick on his holiday? Often these key items can be confirmed at once with a well placed telephone call or a perusal of old records. In one study, in which current histories were compared with old hospital records, it was found that even dramatic events such as hematemesis and melena were often forgotten or fabricated.[11]

Other important information, such as the features of transient neurologic deficits or prior medication use, can often be confirmed by talking with family members or other witnesses. Indeed, patients can improve on their histories by keeping symptom diaries. Such an approach has been shown to substantially decrease the under-reporting of bowel symptoms revealed during routine histories.[12]

Finally, the comparison of present with past diagnostic test results (especially roentgenograms) can do much to eliminate clinical disagreements over the presence, duration and progression of important disease processes.

Confirm key clinical findings with appropriate diagnostic tests: As noninvasive diagnostic tests such as ultrasound and body scanning become more widely available, and as the safety and acceptability of invasive tests such as fibreoptic endoscopy, mediastinoscopy, peritoneoscopy and culdoscopy improve, clinicians can increasingly look to procedures other than autopsy for confirmation of their findings and diagnoses. As we shall see, this strategy of confirmation can be of considerable value in one's continuing education in clinical diagnostic skills as well as in patient care.

Ask "blinded" colleagues to examine your patient: It is often useful to ask a colleague to repeat a key portion of a history or physical to determine whether your observations are confirmed. Such a strategy has been shown to be quite valuable; for example, in the evaluation of patients with suspected valvular heart disease.[13] For this strategy to be useful, however, it is vital that the second examiner be told only of the area to be interrogated or examined and not of your tentative conclusion. Thus, your request should be "Listen to his heart and tell me what you think", rather than "I think this man has aortic stenosis. Please listen and tell me whether you agree." To preface a request for a repeat examination with a statement of your own conclusions downgrades the second examination from a test of clinical agreement to a test of friendship.

"Blind" your assessments of raw diagnostic test data

The objective here is to avoid the effects of prior expectation (remember the tonsils and the fetal heart rates in part I of this series?) by omitting leading information from the history and physical when interpreting raw diagnostic test data. For example, Spodick[14] has emphasized the benefits of interpreting ECGs *twice*, the first time with no information save the patient's age and sex. A similar approach has been advocated by some radiologists as well. Such an approach adds both validity and credibility to the use of terms such as "compatible with" and "diagnostic of".

Report evidence as well as inference

When clinical notes convey the sensory events ("right parasternal heave, loud S1 and P2, faint opening snap, grade 2/4 decrescendo early diastolic murmur with presystolic accentuation best heard at the apex") as well as the corresponding inference ("mitral stenosis"), several benefits accrue. First, agreement between examinations and examiners increases. For example.

*Providing an appropriately private setting for clinical encounters has therapeutic as well as diagnostic implications. For example, Ludy, Gagnon and Caioloa[8] discovered increased patient satisfaction and a doubling of medication compliance when discussions about therapy were held in a private office rather than at the window of a bustling pharmacy.

Table I—Six strategies for preventing or minimizing clinical disagreement

Match the diagnostic environment to the diagnostic task.
Seek corroboration of key findings:
1. Repeat key elements of your examination.
2. Corroborate important findings with documents and witnesses.
3. Confirm key clinical findings with appropriate tests.
4. Ask "blinded" colleagues to examine your patient.

"Blind" your assessments of raw diagnostic test data.
Report evidence as well as inference.
Use appropriate technical aids.
Apply the social sciences, as well as the biologic sciences, of medicine.

Feinstein[15] has documented substantially greater agreement among clinicians for the acoustic elements of cardiac diagnoses than for the diagnoses themselves. Furthermore, we and others[16] have shown that prior discussion and agreement about the pieces of evidence required for a diagnostic inference tend to improve agreement on this diagnostic inference.

Second, the recording of evidence as well as inference often provides crucial information for comparison with later findings, and thus provides a far sounder basis for judging the course and progression of the patient's disease. Third, communication among the multiple clinicians who may be practising "team" care on the same patient is obviously enhanced. Finally, the recording of evidence permits clinicians to retrace their steps back to the initial clinical data when subsequent diagnostic tests or events prove that the original diagnostic inference was wrong.

Use appropriate technical aids

Although the stethoscope is universally accepted as a technical aid to auscultation, the same cannot be said for aids to other important clinical measurements. Measurements of distance and size still rely on fingerbreadths, eggs, fruits and vegetables, despite a consensus that the use of a tape measure or ruler leads to more precise and useful clinical data. Similarly, although the ECG is a routine technical aid for the confirmation and quantitation of cardiac arrhythmias, we often ignore opportunities for the useful application of other technical aids. For example, the "blood pressure" cuff doesn't measure blood pressure at all; it measures air pressure. Thus, it can be used to quantitate abdominal tenderness (quite helpful when following an acute abdomen,[17] especially when sharing clinical responsibilities with other team members), to measure grip strength in arthritis, or as a crude oscillometer in the bedside evaluation of peripheral vascular disease.

Care must be taken not to simply replace human observation with numbers or "hard copy". However, when we can document an important clinical sign by the quick application of an inexpensive technical device that improves the accuracy or precision of clinical measurement we ought to do so.

Apply the social sciences, as well as the biologic sciences, of medicine

The empathetic clinician who takes pains to listen to the patient and to pay close attention to the physician–patient relationship is doing much more than practising the art of medicine; the empathetic clinician is practising good scientific medicine. An understanding of the impact of interpersonal and behavioural factors upon both diagnosis and management is central to the practice of scientific medicine. If we fail to recognize such basic elements as the relation between "not liking" given patients, assigning them bad prognoses and prescribing them multiple drugs,[18] we are at risk of reducing the quality of our clinical care.

Anderson and his colleagues[19] in the department of medicine at King's College Hospital medical school have identified eight skills required to obtain an accurate and useful history and these are listed in Table II. Most of these skills are overtly behavioural, and it is encouraging to discover that the application of empathetic responses to patients can be learned as well as inherited.[20]

This is not to say that the creation of a warm clinician–patient relationship always improves the quality of a medical history. Indeed, one study has documented that patients who most admire and respect their clinicians may withhold key information on the occurrence and severity of the side effects of medication.[21]

Table II—Eight skills required to obtain an accurate and useful history[19]

The ability to:
Establish understanding
Establish information
Interview logically
Listen
Interrupt
Observe nonverbal cues
Establish a good relationship
Interpret the interview

How to learn from one's mistakes

To advance art and science in clinical examination, the equipment a clinician most needs to improve is himself.[7]

We ought to be educated as well as humbled when our patient's subsequent clinical course, operation or autopsy shows us that we erred in our original diagnosis. Indeed, if we were to carry out a more systematic documentation of our own successes and failures in diagnosis we could both identify areas in which we need to improve our clinical skills and find out whether we are making progress in reducing our rate of clinical disagreement. We conclude this clinical epidemiology round with two strategies for documenting and learning from our mistakes.

Both strategies begin by keeping a pocket notebook with a running account of the clinical impressions, conclusions and predictions that one formulates *prior* to the execution of a confirmatory procedure (such as a biopsy or other definitive diagnostic test, an operation or an autopsy). These clinical impressions or predictions can then be compared with the results of the confirmatory procedures in one of two ways.

The method of simple agreement

This method is illustrated in Table III, where, for example, patient A.B.C. was judged to have atrial fibrillation on the basis of a history of palpitations and the finding of an irregularly irregular pulse; this was confirmed on a subsequent electrocardiogram. Similarly, the impressions of breast cancer in patient G.H.I. and anemia in patient J.K.L. were confirmed by subsequent definitive procedures. However, the pleural effusion in patient D.E.F. was missed, and the impression of hyperthyroidism in patient M.N.O. was not confirmed by serum T_4 and T_3 determinations. Thus, the rate of simple agreement was three out of five (60%).

The method of simple agreement

does tell us how we are doing in general terms and, if compared with a similar account kept previously, could give a rough idea of whether we are making progress in reducing this type of clinical disagreement. Furthermore, this method could point out areas in which we might want to brush up our clinical skills (in this case, in the physical examination of the chest and in assessing the signs of hyperthyroidism).

However, as Shapiro[22] has pointed out, the method of simple agreement has a major drawback: it treats clinical impressions as "all-or-none" judgements and fails to recognize that they are, with rare exceptions, contemplated in terms of probabilities (patient M.N.O. "*may* have hyperthyroidism" and patient A.B.C. "*almost certainly* has atrial fibrillation"). For this reason Shapiro proposed a more complex but much more useful "predictive accuracy" method.

The method of predictive accuracy

This method is illustrated in Table IV, where the clinical impressions and predictions from Table III reappear, but this time the clinician has also entered the probability with which he believes his impression will be confirmed by subsequent, definitive studies. Because this method requires that clinicians convert words such as "probably", "almost certainly" and "clear-cut" to numbers, some will find its implementation, at least initially, to be both cumbersome and disconcerting. At any rate, our intrepid clinician in Table IV judges

Table III—The method of simple agreement for assessing clinical impressions and predictions

	Clinical impression or prediction		Confirmation		Agreement*	
Patient	Evidence	Inference	Procedure	Result	Yes	No
A.B.C.	History of palpitations; irregularly irregular pulse	Atrial fibrillation	Electrocardiogram	Atrial fibrillation	X	
D.E.F.	Percussion and auscultation normal	Normal chest examination	Chest x-ray	Left pleural effusion		X
G.H.I.	Hard 2-cm lump in left breast; negative axillae	Breast cancer	Biopsy and surgical specimen	Breast cancer	X	
J.K.L.	History of fatigue; pale mucosae and nail beds	Anemia	Hemoglobin determination	Low	X	
M.N.O.	Anxious-looking, sweaty; fine tremor at rest	Hyperthyroid	Measurement of serum T_4 and T_3	Normal		X

*Simple agreement: 3/5 (60%).

Table IV—The method of predictive accuracy for assessing clinical impressions and predictions

	Clinical impression or prediction			Confirmation			
Patient	Evidence	Inference	Probability	Procedure	Result	Probability at confirmation	Predictive accuracy*
A.B.C.	History of palpitations; irregularly irregular pulse	Atrial fibrillation	0.95	Electrocardiogram	Atrial fibrillation	0.95	0.93
D.E.F.	Percussion and auscultation normal	Normal chest examination	0.67	Chest x-ray	Left pleural effusion	1 − 0.67 = 0.33	−0.60
G.H.I.	Hard 2-cm lump in left breast; negative axillae	Breast cancer	0.75	Biopsy and surgical specimen	Breast cancer	0.75	0.58
J.K.L.	History of fatigue; pale mucosae and nail beds	Anemia	0.80	Hemoglobin determination	Low	0.80	0.68
M.N.O.	Anxious-looking, sweaty; fine tremor at rest	Hyperthyroid	0.50	Measurement of serum T_4 and T_3	Normal	1 − 0.50 = 0.50	0

*Average predictive accuracy = total predictive accuracy/number of cases = 1.59/5 = 0.32 (accuracy coefficient).

the atrial fibrillation in patient A.B.C. to be 95% likely to be confirmed by the subsequent ECG, believes that the odds are two to one that patient D.E.F. will have a normal chest x-ray, thinks the chances are three out of four that patient G.H.I. has breast cancer, and so on.

When the subsequent confirmatory tests show that the earlier prediction was correct the corresponding probability value appears in the column titled "Probability at confirmation"; this has been done for patients A.B.C., G.H.I. and J.K.L. However, when the earlier clinical prediction is subsequently shown to be wrong, the entry in this column becomes the predicted probability subtracted from 1.0; thus, the entry for patient D.E.F., who did not have a normal chest x-ray, becomes 1.0 − 0.67 = 0.33, and the entry for patient M.N.O., who did not have hyperthyroidism, becomes 1.0 − 0.50 = 0.50. Notice from the last column in Table IV and from the first entry in Table V that a probability of 0.50 contributes nothing to predictive accuracy, right or wrong; there is no reward for indifference.

The next step can either be accepted on faith or studied in more detail elsewhere.[22] In brief, one rewards correct diagnoses and rewards those predicted with high probability (e.g., for patient A.B.C.) more than those predicted with lower probability (e.g., for patient G.H.I.). As it happens, one good way to do this is to use a strategy based on "information theory"; the resulting scores appear in the last column of Table IV, titled "Predictive accuracy". Since the calculations of predictive accuracy are cumbersome, we have summarized them in Table V for those probabilities likely to be used with greatest frequency in clinical settings.

Finally, to find out how one is doing overall, one can simply sum the numbers in the predictive accuracy column of Table IV and divide this total by the number of entries. In this example the resulting average predictive accuracy (which Shapiro[22] has named the accuracy coefficient) is 0.32, a value that compares favourably with that of seasoned clinicians predicting, for example, whether given patients with systemic lupus erythematosus will develop the nephrotic syndrome.*

Systems such as these for identifying and quantifying one's own clinical disagreements are not widely used in medicine, but we hope that one result of this series of clinical epidemiology rounds will be their application and evaluation on a much wider scale. Furthermore, we hope that readers of these rounds will report back to us whether these rounds are useful in the "front lines".

Table V—Predictive accuracy scores for various prediction probabilities

Prediction probability	Predictive accuracy score*: When confirmed	Predictive accuracy score*: When not confirmed
0.5	0	0
0.6	0.26	−0.32
0.67	0.42	−0.60
0.7	0.49	−0.74
0.75	0.58	−1.0
0.8	0.68	−1.32
0.9	0.85	−2.32
0.95	0.93	−3.32

*Based on the formulas: $\log_2$ (prediction probability) + 1 if confirmed; and $\log_2$ (1 − prediction probability) + 1 if not confirmed.

*The devious reader will have noted that clinicians can artificially inflate their accuracy coefficients by firmly asserting (with prediction probability = 0.95) that each of their patients does *not* have canine distemper; such strong convictions about the absence of impossible diagnoses will help the score but not the patient, and are useless in one's continuing education.

References

1. Steen PA, Tinder JH, Tarhan S: Myocardial infarction after anesthesia and surgery. *JAMA* 239: 2566, 1978
2. Stollerman GH, Markovitz M, Taranta A, et al: Jones criteria (revised) for guidance in the diagnosis of rheumatic fever. *Circulation* 32: 664, 1965
3. Hull R, Delmore T, Genton E, et al: Warfarin sodium versus low-dose heparin in the long-term treatment of venous thrombosis. *N Engl J Med* 301: 855, 1979
4. Feinstein AR: Scientific methodology in clinical medicine. I. Introduction, principles and concepts. *Ann Intern Med* 61: 564, 1964
5. Idem: Scientific methodology in clinical medicine. II. Classification of human disease by clinical behavior. Ibid, p 757
6. Idem: Scientific methodology in clinical medicine. III. The evaluation of therapeutic response. Ibid, p 944
7. Idem: Scientific methodology in clinical medicine. IV. Acquisition of clinical data. Ibid, p 1162
8. Ludy J, Gagnon J, Caioloa S: The patient–pharmacist interaction in two ambulatory settings — its relationship to patient satisfaction and drug misuse. *Drug Intell Clin Pharm* 11: 81, 1977
9. Sackett DL: Clinical diagnosis and the clinical laboratory. *Clin Invest Med* 1: 37, 1978
10. Yerushalmy J, Harkness JT, Cope JH, et al: The role of dual reading in mass radiography. *Am Rev Tuberc* 4: 443, 1950
11. Corwin RG, Krober M, Roth HP: Patients' accuracy in reporting their past medical history, a study of 90 patients with peptic ulcer. *J Chronic Dis* 23: 875, 1971
12. Manning AP, Wyman JB, Heaton KW: How trustworthy are bowel histories? Comparison of recalled and recorded information. *Br Med J* 2: 213, 1976
13. Feinstein AR, Di Massa R: The unheard diastolic murmur in acute rheumatic fever. *N Engl J Med* 260: 1331, 1959
14. Spodick DH: On experts and expertise: the effect of variability in observer performance. *Am J Cardiol* 36: 592, 1975
15. Feinstein AR: *Clinical Judgment*, Williams & Wilkins, Baltimore, 1967, pp 322-27
16. McLachlan MSF, Thomson JG, Taylor DW, et al: Observer variation in the interpretation of lower limb venograms. *Am J Roentgenol* 132: 227, 1977
17. Shafer N: Technique for quantitating abdominal pain. *JAMA* 201: 558, 1967
18. Ehrlich HJ, Bauer ML: Therapists' feelings toward patients and patient treatment and outcome. *Soc Sci Med* 1: 283, 1967
19. Anderson J, Day DL, Dowling MAC, et al: The definition and evaluation of the skills required to obtain a patient's history of illness: the use of videotape recordings. *Postgrad Med J* 46: 606, 1970
20. Fine VK, Therrien ME: Empathy in the doctor–patient relationship: skill training for medical students. *J Med Educ* 52: 757, 1977
21. Snyder D, Lynch JJ, Gruss L: Doctor–patient communications in a private family practice. *J Fam Pract* 3: 271, 1976
22. Shapiro AR: The evaluation of clinical predictions. A method and initial application. *N Engl J Med* 296: 1509, 1977