Does specialty board certification influence clinical outcomes?

Eric N. Grosch MD
United Urgent Care Clinic, Fort Myers, FL, USA

Correspondence
Eric N. Grosch
United Urgent Care Clinic,
3594 Broadway Suit A,
Fort Myers, FL 33901, USA
E-mail: eric.grosch@gmail.com

Keywords: board certification, causal criteria, data dredging, outcomes, syllogism

Accepted for publication:
2 September 2004

Abstract

Background The public seems to crave a simplistic index of 'quality', analogous to 'The Good Housekeeping Seal of Approval', for the complex endeavour of clinical medicine. The American Board of Medical Specialties (ABMS) and its member boards have purported to fill the vacuum in an effort that bears many of the earmarks of a public relations publicity campaign. The author examined the validity of the evidence touted in support of that effort. **Methods** By applying Hill's causal epidemiologic criteria and logical and statistical inference, the author evaluated: (i) published data sources consisting of the most comprehensive collection of studies yet gathered that purports to provide evidence of the relevance of board certification to clinical outcomes, a collection presented by Sharp *et al.* apparently with the advice and consent of ABMS, that they posited as containing 'relevant findings', to what purpose they left unspecified; and (ii) the review article of Sharp *et al*. **Results** The data that Sharp *et al*. presented provided no credible link between specialty board certification and outcomes or 'quality' of clinical care. Sharp *et al*. ignored the evidence of absent evidence they found and proposed enthusiastic but unjustified conclusions in support of specialty board certification as an index of clinical 'expertise'. **Conclusions** No evidence supports the touted clinical benefit of specialty board certification. Specialists in clinical medicine and surgery are unamenable to simplistic evaluation by examination, yet specialty board certification remains an *ersatz* standard of doctors' clinical quality in the absence of supporting evidence.

The review article of Sharp *et al*. (2002) is the only comprehensive survey of studies purporting to link board certification with clinical outcomes and it appears to be part of a campaign of the American Board of Medical Specialties (ABMS), the study's only named supporter, to legitimize the notion that doctors who hold board certification produce better clinical outcomes than others.

A clinician acquires clinical acumen from education and training, which within any specialty is much the same, not from passing a multiple choice quiz.

A doctor may gain knowledge by preparing for a board certification examination, so, based on the notion that board-certified practitioners know more than non-certified practitioners, clinical care outcomes should improve accordingly. This seems to be the rationale for the subject link. A seeming reasonable construction of the claimed link is that board certification *causes* improvement in clinical care outcome, a link that seems amenable to evaluation by Hill's causal criteria (Hill 1965), common sense cause-and-effect logical triangulations of relations in complex systems, such as biological systems, for which detailed elucidation of mechanisms is unfeasible:

Hill's Causal Criteria

I Strength. A strong association provides firmer evidence of causality than a weak one.

II Consistency. The greater are the number and diversity of methods of study, populations, circumstances, investigations and cogent studies that show a similar, consistent association, the stronger is the causal evidence.

III Specificity. Only the putative cause should lead to change by some measure. (Just as it is improbable that epidemiologists can demonstrate that risk factors have consistent influence on most chronic diseases, so it is improbable that investigators will ever be able to demonstrate that board certification has a salutary influence on clinical outcomes.)

IV Temporality. Exposure to the putative cause should precede the change; 'the cart does not precede the horse' (E leads to D, not D leads to E).

V Biologic gradient. An increase in the level, intensity, duration, or total level of exposure to the putative cause leads to progressive increases in the magnitude of effect on the outcome. This is in keeping with the dose–response relationship of biologic phenomena. Thus, board certification test scores should correlate with quality of clinical care outcomes.

VI Plausibility. It is helpful for a plausible, well defined explanation based on a known fact or model to underlie a consistent association.

VII Coherence. All available evidence of the change, in relation to the putative cause should 'stick together' (cohere) to form a cohesive whole. The proposed causal relationship should not conflict with information from other knowledge sources.

VIII Experimentation. Clinical trials may yield evidence in support of a causal hypothesis.

IX Analogy. Analogy is weak evidence but it implies a similarity in some respects among things that are otherwise different.

The more criteria each primary study satisfies, the more plausible is the link and vice versa. In the following reckoning, beneath evaluation of each primary study, the foregoing Roman numerals designate the causal criteria that each failed to satisfy.

Circular Reasoning (X)

Generalizing the link between knowing more and improved outcome of treatment for one condition, if shown, would, by causal criteria II/III, require confirmation that each board-certified practitioner's clinical outcomes are consistently superior to those of a non-certified practitioner's, at least for the most common clinical conditions.

A syllogism, analogous to the following classic syllogism, completes the thought:

All men are mortal.
Socrates is a man.
Therefore, Socrates is mortal.

For this application:

Each board certified practitioner knows more than each non-certified practitioner.

Each practitioner in the study whose clinical outcomes for the condition of interest are superior to those for each non-certified practitioner in the study is a board-certified practitioner.

Therefore, each practitioner in the study whose clinical outcome for the condition of interest is superior to those for each non-certified practitioner knows more than each non-certified practitioner in the study.

Leaping, on faith alone, from particular to general:

Each practitioner in the study whose clinical outcome for the condition of interest is better than that for each non-certified practitioner is representative of every board-certified practitioner.

Therefore, each board-certified practitioner knows more than each non-certified practitioner.

The last proposition, which the study sought to prove, is identical to the major premise of the syllogism. Using a conclusion as a premise of an argument is to perpetrate the fallacy of *petitio principii* (arguing in a circle, begging the question) (Ingle 1972), designated below by Roman numeral X.

Data Dredging (XI)

Sackett cited biases, including:

(b) Data dredging bias (looking for the pony). When data are reviewed for all possible associations without prior hypothesis, the results are suitable for hypothesis-forming activities only (Sackett 1979, p. 62).

Retrospectivity of a study, by necessary preclusion of prospective study design, *ipso facto,* entails data dredging and its findings apply to hypothesis forming, not conclusions.

Some patients under some practitioners' care may yield superior measures of clinical outcome, such as laboratory tests (Ramsey *et al.* 1989). Concluding that one such laboratory-test result by itself designates the practitioner whose care produced that outcome as clinically superior, in general, constitutes data-dredging. Sharp *et al*. and the authors of their references applied that *modus operandi* repeatedly. Roman numeral XI designates studies that rely on data dredging.

Spurious Statistical Significance (XII)

Wood (1977, p. 305) elucidated the distinction between statistical and clinical significance with an example, table 1, below.

Table 1 Comparison of outcomes for two treatment groups

		Outcome		
		A	*B*	*Total*
Treatment	1	4	4	8
	2	8	24	32

Outcome figures appear to show treatment 2 to be superior to treatment 1 in producing outcomes A and B but most effective in producing outcome B. The outcomes were A, male calves; B, female calves. The treatment groups, 1 and 2, were mother cows, whose pregnancies had developed from artificial insemination, administered in two ways at insemination: one group faced north and the other south. Because no known mechanism of sex determination of calves relies on the direction a prospective mother faces when undergoing artificial insemination, the results, although 'statistically significant', are unlikely to recur in subsequent trials, so they are clinically insignificant.

Another source of spurious statistical significance is too large a sample population:

> (i) Wrong sample size bias. Samples which are too small can prove nothing; samples which are too large can prove anything (Sackett 1979, p. 61).

'Hill's causal criterion VI provides a conceptual standard to evaluate the plausibility, and criterion VII, the coherence, together, the credibility, of the explanatory model of a phenomenon of interest. The phenomenon of interest here is the proposed link between board certification and beneficial outcomes of clinical care. A recurrent impediment to the plausibility and coherence among the studies reviewed is the confusion of spurious statistical significance for clinical significance. Hill's causal criteria VI and VII here highlight the spurious nature of those instances of statistical significance and Roman numeral XII flags them where they occur.'

Sharp *et al*. described scale-degradation bias:

Scale-Degradation Bias (XIII)

> . . . most studies pooled patient data across physicians, negating the possibility of measuring the individual physician's performance (Sharp *et al*. 2002, p. 30).
>
> (c) Scale degradation bias. The degradation and collapsing of measurement scales tends to obscure differences between groups under comparison (Sackett 1979, p. 62).

Roman numeral XIII designates studies that feature scale-degradation bias.

Superfluous Significance Testing (XIV)

Cruess (1988) referred to Homeida *et al*. (1988) in citing superfluous significance testing of obvious differences, an error designated by Roman numeral XIV.

Article Review

Sharp *et al*. examined only retrospective studies, except for that of Ramsey *et al.* (1989) which mixed prospective testing of board-certified and non-certified internists with part of a board-certification examination in internal medicine (X) and retrospective comparison of their clinical outcomes. They claimed:

> . . . higher scores on certification examinations correlate with measures of better patient care (Sharp *et al*. 2002, p. 535).

Ramsey, former chairman of the American Board of Internal Medicine (ABIM), could correlate certification neither with improved clinical care outcomes nor increased medical knowledge (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII).

Norcini *et al*. (1985) characterized scores that candidates achieved on the multiple choice quiz that the ABIM used to determine board certification as indicative of their 'medical knowledge,' without having produced objective evidence that board-certification was indicative thereof, then compared the 'medical knowledge' (scores on the quiz) of board-certified and non-certified doctors to demonstrate that board-certified doctors had more 'medical knowledge.' Norcini *et al.* (1985) thereby committed the fallacy of *petitio principii* (circular reasoning) (X).

Norcini *et al*. (1987) omitted consideration of clinical outcomes by linking only tested knowledge with ratings from training and conceded that the 'halo effect' may confound ratings. Sharp *et al*. claimed 'ratings in training correlate with clinical knowledge' (I, II, III, IV, V, VI, VII, VIII, IX, XI).

Sharp *et al*. discounted Ferry *et al.* (1985), Greenfield *et al*. (1995), Nelson *et al*. (1993), doubted the reliability of Sloan *et al*. (1989) and Tussing & Wojtowicz (1993) and based their conclusions on Adamson *et al*. (1975), Brook *et al*. (1990), Haas *et al.* (1995), Heck *et al.* (1998), Kelly & Helliger (1986, 1987), Morrison (1998), Nelsen *et al*. (1994), Norcini *et al*. (2000), Pearce *et al*. (1999) and Schwartz & Mendelson (1989):

Tussing and Wojtowycz studied 65 784 of the 1986 deliveries in New York but examined no clinical records to determine motivations for Caesarean section.

Tussing and Wojtowycz had no non-certified obstetricians (OB) to compare with board-certified OBs, so they compared groups known, *a priori*, to differ: board-certified OB specialists and non-OB specialists, unspecified, possibly general practitioners (GPs), family practitioners (FPs) or others, who may or may not have been board certified in their specialties:

> ...Tussing and Wojtowycz found that board-certified obstetricians had a higher cesarean-section rate than did a group of physicians from a mix of specialties. Because the study did not adjust sufficiently for case mix, this finding may be attributable to the fact that the certified obstetricians dealt with more high-risk pregnancies (Sharp *et al*. 2002, p. 536).

Sharp *et al*. neglected to consider other possible explanations for that difference and had insufficient data from Tussing and Wojtowycz, to verify that or any other explanation.

Sharp *et al*. agreed with Tussing and Wojtowycz in claiming that the difference in Caesarean rate was statistically significant at the $P < 0.1\%$ level ($P < 0.001$ level). For reasons best known to themselves, Sharp *et al*. suppressed the clinically insignificant difference between Caesarean rates: 28.6% for OBs and 27.8% for all doctors (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV).

Sloan *et al*. found a malpractice rate in favour of non-certified doctors, contrary to expectation. Sharp *et al*. slanted their comments in favour of board certification:

> The certified physicians could have cared for more complicated patient populations than their non-certified colleagues and generated more malpractice claims.

Sharp *et al*. considered no other explanations for 'more malpractice claims' and had insufficient data from Sloan *et al*. to verify that or any other explanation.

Consistent with their bias, Sharp *et al*. excluded Sloan *et al*. from consideration and, as justification, reasonably asserted, 'malpractice claims do not necessarily reflect inferior quality of care', a proposition that was true and known initially, so, assuming that 'quality of care' depends on 'outcomes', the reason that the study of Sloan *et al*. survived the winnowing process of Sharp *et al*. in the first place seems obscure (I, II, III, IV, V, VI, VII, VIII, IX, XI, XII, XIII).

> ...Heck and colleagues compared board-certified and non-certified orthopedic practitioners' performances on knee replacements... found no association with certification status. The study was limited by the fact that 41 practitioners were board-certified, compared with only seven who were not (Sharp *et al*. 2002, p. 536).

Heck *et al*. provided Sharp *et al*. no evidentiary basis for distinguishing outcomes between patients of board-certified and non-certified orthopaedists. Pooled data removed any possibility of proving the subject link:

> In 1989 there were 12 425 orthopaedic surgeons who had received certification under the auspices of the American Board of Medical Specialties. According to the American Medical

> Association, there were 18 741 self-declared orthopaedists in the United States . . . at the time of commencement of the PORT activities, the percentage of Board Certified Orthopaedists in the United States was 66%, . . . 85% of the participating surgeons in this series successfully had completed certification by the American Board of Medical Specialties. Although the impact of this certification process on the patient outcomes in this sample was not shown, the sample may be biased toward better patient outcomes on this basis (Heck *et al.* 1998, p. 98).

Improvement in technology from 1989 to 1998 may also have yielded 'better patient outcomes' (I, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII).

Kelly & Hellinger (1986) stated, from their table 4:

> Stomach operation w/ ulcer diagnosis
>
> | | Coefficient | (t-ratio) |
> |---|---|---|
> | Board certified in surgery | –0.0207* | (1.646) . . . |
>
> *indicates the result is significant at the 0.10 level, using a two-tailed test (p. 794).
>
> Surgical patients with ulcers are 2% less likely to die in the hospital when their primary surgeon is board certified in surgery, compared with patient of non-board-certified physicians (p. 796).

Table 4 purports to show that the difference in mortality for board-certified and non-certified surgeons in peptic ulcer disease is statistically significant but, to exclude a spurious link, consistent with Hill's causal criteria VI/VII, the authors should explain the advantages that a board-certified surgeon has but that a non-certified surgeon does not have that apply to the mortality rate of gastric surgery for peptic ulcer disease without also applying to any other procedure, especially in view of the greater hazard of at least one [abdominal aortic aneurysmectomy (AAA)], compared to peptic ulcer surgery. The risk that the indicated difference in mortality could occur by chance alone is arguably too high at 0.10, a figure that Sharp *et al*. suppressed. Even if that *P* value designated a statistically significant difference, the two percent mortality difference is clinically insignificant (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII).

In Kelly and Hellinger, a retrospective study displaying no original data, hence an uninterpretable one, table 5 tabulated no figures for family practitioners (FPs) for cardiac catheterization. For myocardial infarction, table 6 tabulates an unexplained 'coefficient' and '*t*-ratio', said to be:

> . . . statistically significant at the 0.05 level, using a two-tailed t-test (Kelly & Hellinger 1987, p. 382).

Sharp *et al*. did not say whether the differences in mortality among patients from heart attack between certified and non-certified internists and FPs were clinically significant.

Kelly and Hellinger discussed assumptions of others:

> . . . because investigators could not link physician and patient data, they had to assume that each physician on a hospital's medical staff treated the average number of patients . . . They also found that physician board certification is not related to the mortality of AMI patients, when board certification status is measured by the percent family practitioners and internists on staff who are board certified (Kelly & Hellinger 1987, p. 389).

Trying to justify the efficacy of board certification, ostensibly a recognition of global medical knowledge, by considering differences in outcomes of only one or two conditions is invalid by Hill's causal criteria II/VI/VII. Sharp *et al*. did not say and had insufficient data to investigate whether the differences in outcomes of other clinical conditions between certified and non-certified internists and FPs were statistically or clinically significant and provided no mechanisms for any proposed influence of board certification of treating doctors on outcomes of any clinical conditions (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII).

Pearce *et al*. (1999) compared obviously dissimilar groups:

> board-certified surgeons with subspecialty certification in vascular surgery from the American Board of Surgery with non-certified general surgeons . . .

Subspecialty certification in vascular surgery necessitates also subspecialty training in vascular surgery. In a quest to affirm board certification alone, as determinative in improving clinical outcome (III), the necessary concomitant of salient training acts as a 'confounder'.

One would expect better outcomes from surgery by subspecialists than by other surgeons, so testing statistical significance (*P* values) of differences in outcomes expected, *a priori*, is superfluous: board-certified vascular surgeons produced a 15% lower risk of death or complications (proportion of each, unspecified), after carotid endarterectomy ($P = 0.002$), and a 24% lower risk of death after AAA ($P = 0.009$), than non-certified surgeons in other surgical specialties, cardiac surgery, neurosurgery and general surgery.

In the discussion following the article, Dr Pearce admitted the seemingly anomalous finding that board-certified vascular surgeons had a higher rate of postoperative amputations for femoral–popliteal bypass than did the other surgical specialists but did not attempt to explain it. Pearce *et al*. provided insufficient data to verify any explanation of that discrepancy and Sharp *et al*. omitted any mention of it (I, II, III, IV, V, VI, VII, VIII, IX, XI, XII, XIII, XIV).

After having produced no empirical evidence, in support of board certification, Sharp *et al*. ignored the evidence they had adduced and asserted:

> . . . empirical evidence supports the value of board certification (Sharp *et al*. 2002, p. 537).

They went on to assert:

> Certification has been associated with increased medical knowledge (Ramsey *et al.* 1989), superior training (Norcini *et al*. 1987), and certain aspects of patient care (Ramsey *et al.* 1989, Haas *et al.* 1995, Pearce *et al*. 1999), Sharp *et al*. (2002, p. 537).

Pearce *et al*. (1999) correlated the extent of surgical experience in training ('surgeon volume and training') with clinical outcomes. To qualify for board certification in vascular surgery, a candidate must not only present documentation of experience in training to that board but he must also pass a multiple choice quiz, the feature that distinguishes board certification from experience (in training) alone. The evidence for that assertion is the fee he must pay to the board to sit and take the quiz. The board does not grant him certification without having verified his passing score on the quiz and having received the fee from him. The candidate pays no fee for the experience he acquires in training or for its documentation. A non-certified vascular surgeon, then, has experience in training equivalent to that of a board-certified vascular surgeon. Pearce *et al*. did not therefore correlate board certification *per se*, that is, with respect to its features (quiz, fee), separately from confounding experience in training, with clinical outcomes (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII).

Haas *et al.* (1995) pooled board-certified obstetricians (OBs) and FPs, compared an inadequate number of non-certified FPs and showed equivocal and overlapping figures in linking clinical care outcomes to doctors' board-certification status (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII).

Nelsen *et al*. (1994) compared board-certified and non-certified FPs' rates of procedural usage, an issue irrelevant to clinical outcomes (I, II, III, IV, V, VI, VII, VIII, IX, XI, XIII).

Adamson (1975), Brook *et al*. (1990), Morrison (1998), Nelsen *et al*. (1994) and Schwartz & Mendelson (1989) conducted retrospective studies of large, pooled databases, which are inconclusive (I, II, III, IV, V, VI, VII, VIII, IX, XI, XIII).

Sharp *et al*. (2002) attributed 'sound design' (p. 538) to Norcini *et al*. (2000), although it suffered from many deficits aforementioned, pooled data and data dredging of retrospective analysis, among others (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XIII, XIV).

Carver (1978) described the proper scientific method as first determining whether results are scientifically significant, then whether they are statistically significant. In Carver's 'corrupt scientific method', the investigator determines whether results are statistically significant first and, only if they are, does he pursue a determination of whether they are scientifically significant. In every instance, above, Sharp *et al*. have insufficient data to pursue the proper scientific method, so they can rely only on the corrupt scientific method.

Sharp *et al*. perpetrated *non sequiturs* (Ingle 1972, p. 263).

> In this era of evidence-based medicine, clinical outcomes have become the 'gold standard' for evaluating the quality of care (Sharp *et al*. 2002, p. 537).

If clinical outcomes were the 'gold standard', letters of reference would include doctors' clinical outcomes but they do not (Keim *et al*. 1999) and credentialers would inquire about them but they fixate instead on board-certification status (JCAH 1971; Lang 1999; Federation of State Medical Boards of the United States 2000; JCAHO 2003):

> . . . Those who do not know a poet when they see one do not weigh his achievements but his credentials . . . (Murphy 1981, p. 28).

Those who do not know a doctor when they see one do the same.

Sharp *et al*. confirmed their misunderstanding about the 'gold standard', outcomes, in their assertion about hypothetical, not current, prospects for 'valid outcome measures' (p. 537).

> Selecting or developing valid outcome measures of practice performance for specialty boards' databases would make it feasible to examine the relationships between board certification and patients' outcomes . . . (Sharp *et al*. 2002, p. 541).

Even assuming the validity of outcome measures, quiz scores, the primary data that purveyors of board certification seek to correlate with outcome measures, are themselves wanting:

> . . . [T]he traditional examination system . . . achieves . . . pseudoprecision, for it has chosen the accurate measurement of the barely relevant in preference to the less precise measurement of the most highly relevant . . . [N]umerical ascriptions and grades, often of very dubious reliability and validity . . . added together and averaged . . . produces results . . . of little or no predictive validity . . . (Simpson 1976, p. 22).

Sharp *et al*. pointed out that 'the public' errs:

> The public also uses board certification as a measure of a physician's expertise, despite well-documented statements by the ABMS and the member boards that board certification is but one of several qualifications to be considered in assessing the quality of a physician's clinical care (Sharp *et al*. 2002, p. 534).

Despite that reasonable declaration, Sharp *et al*. asserted:

> Despite the lack of unequivocal evidence documenting the value of board certification, we do not advocate removing it as a measure of expertise (Sharp *et al*. 2002, p. 541).

Sharp *et al*. thus ignored the evidence that they had adduced against board certification but applied:

> (a) The biases of rhetoric. Any of several techniques used to convince the reader without appealing to reason (Sackett 1979, p. 60).

Authors of a scientific article have an ethical duty to advocate whatever conclusion their evidence supports. Sharp *et al*. had adduced no unequivocal evidence to justify using board certification as a measure of expertise, so, unless they considered equivocal evidence conclusive, which would be a contradiction in terms, removing board certification as a measure of expertise was their only appropriate recommendation. If authors of scientific articles failed consistently thus to ignore evidence, scientific literature would lose utility.

Daniel Lang, self-proclaimed expert in doctor credentialing, generalized the point that indeterminately many official credentialers share the folly of 'the public':

> Some organizations would not even grant an interview to a physician who is not board certified, let alone one with a history of one or more failed attempts. Managed care organizations, particularly . . . use the board certification requirement for making the first cut. The use of flags . . . will become increasingly arbitrary and inflexible as the physician glut increases (Lang 1999, p. 3).
>
> . . . in a buyer's market there is a tendency to overvalue the file without blemish as a predictor of the future organization-physician relationship . . . (Lang 1999, pp. 2, 3, 12).

After touting 'high standards' for doctors, expert Lang endorsed dispensing with doctors in favour of non-doctors and indeterminate others:

> . . . Countless articles contend that traditional specialty activities can be safely and effectively performed by nonphysicians and other health care providers . . . (Lang 1999, p. 184).

As *ersatz* for 'countless articles', expert Lang cited two. Cost containment appears to trump 'quality' of 'qualifications' in 'health care' decision making.

Sharp *et al*. continued:

> Perhaps one lesson to be learned from this review is the need to thoughtfully examine this recertification process to document its value and assure the American public that continued certification is a marker of high-quality care (Sharp *et al*. 2002, p. 541).

Doubtless, '[o]ne lesson to be learned' is the one Sharp *et al*. cited but, if they had learned the lesson they cited, they should have postponed submission of their article until they had performed said thoughtful examination of their evidence, whereupon, that thoughtful mood must rationally have compelled them to conclusions contrary to those they have espoused.

Evidence-based medicine is a conceptual framework that purports to inform the correctness of answers to the board certification examination, so for those attempting to validate board certification itself to exempt their assessment from the precepts of evidence-based medicine seems inconsistent.

Documentation of the value of certification and recertification does not exist, as Sharp *et al*. conceded, so assuring the American public that 'continued certification' be 'a marker of high quality care' would constitute fraud. In this era of evidence-based medicine, absence of an evidentiary basis for such important social conduct as doctor selection seems ethically dubious.

> It can scarcely be questioned that when truth or falsehood of an event or observation may have important bearings on conduct, over-doubt is more socially valuable than over-credulity (Pearson 1911, p. 51).

Many health care organizations now disregard Pearson's counsel to the extent that their over-credulity in the value of board certification bears heavily on their social conduct in excluding doctors who do not hold it (Lange, *supra*). Accordingly, doctors feel undue compulsion to obtain board certification to earn their livelihoods. The ABMS and its member boards profit from such over-credulity, yet ABMS and its member boards have declared no moratorium on board certification to redress that ethical conflict. That collective omission of appropriate social action arguably leaves to the doctor himself the ethical duty to desist from seeking board certification, lest he support such ethically dubious practice.

References

Adamson J.E. (1975) Five-year history of the American Society of Plastic and Reconstructive Surgeons, 1969–1973. *Plastic and Reconstructive Surgery* **55** (4), 445–559.

Brook R.H., Park R.E., Chassin M.R., Kosecoff J., Keesey J. & Solomon D.H. (1990) Carotid endarterectomy for elderly patients; predicting complications. *Annals of Internal Medicine* **113** (10), 747–5312.

Carver R.P. (1978) The case against statistical significance testing. *Harvard Educational Review* **48** (3), 378–9921.

Cruess D.F. (1988) Review of the use of statistics in the *American Journal of Tropical Medicine and Hygiene* for January–December. *American Journal of Tropical Medical and Hygiene* **41** (6), 619–626.

Federation of State Medical Boards of the United States. (2000) *A Guide to the Essentials of a Modern Medical Practice Act.* 9th edn. April http://www.fsmb.org/Policy%20Documents%20and%20White%20Papers/9th_edition_essentials.htm.

Ferry M.E., Lamy P.P. & Becker L.A. (1985) Physician's knowledge of prescribing for the elderly. A study of primary care physicians in Pennsylvania. *Journal of the American Geriatric Society* **33** (9), 616–625.

Greenfield S., Rogers W., Mangotich M., Carney M.F. & Tarlov A.R. (1995) Outcomes of patients with hypertension and non-insulin dependent diabetes mellitus treated by different systems and specialties: results from the medical outcomes study. *JAMA* **274** (18), 1436–1444.

Haas J.S., Orav E.J. & Goldman L. (1995) The relationship between physicians' qualifications and experience and the adequacy of prenatal care and low birthweight. *American Journal of Public Health* **85** (8), 1087–1091.

Heck D.A., Robinson R.L., Partridge C.M., Lubitz R.M. & Freund D.A. (1998) Patient outcomes after knee replacement. *Clinical Orthopaedics and Related Research* **356**, 93–110.

Hill A.B. (1965) The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine* **58** (5), 295–300.

Homeida M., Abdel-Gadir A.F., Cheever A.W., Bennett J.L., Arbab B.M., Ibrahim S.Z., Abdel-Salam I.M., Dafalla A.A. & Nash T.E. (1988) Diagnosis of pathologically confirmed Symmers' periportal fibrosis by ultrasonography: a prospective blinded study. *American Journal of Tropical Medicine and Hygiene* **38** (1), 86–91.

Ingle D.J. (1972) Fallacies and errors in the wonderlands of biology, medicine and Lewis Carroll. *Perspectives in Biology and Medicine* **15** (2), 254–281, 263, 273.

Joint Commission on Accreditation of Health Care Organizations (JCAHO). (2003) *Comprehensive Accreditation Manual for Hospitals*. JCAHCO, Chicago.

Joint Commission on Accreditation of Hospitals (JCAH). (1971) *Guidelines for the Formulation of Medical Staff Bylaws, Rules and Regulations.* JCAH, Oakbrook Terrace.

Keim S., Rein J.A., Chisolm C. *et al.* (1999) Standardized letter of recommendation for residency application. *Academic Emergency Medicine* **6** (11), 1141–1146.

Kelly J.V. & Hellinger F.J. (1986) Physician and hospital factors associated with mortality of surgical patients. *Medical Care* **24** (9), 785–800.

Kelly J.V. & Hellinger F.J. (1987) Heart disease and hospital deaths: an empirical study. *Health Services Research* **22** (3), 369–395.

Lang D.A. (1999) *Medical Staff Peer Review: Motivation and Performance in the Era of Managed Care.* American Hospital Publications, Health Forum, Chicago.

Morrison W.P. (1998) Physicians disciplined by a state medical board. *JAMA* **279** (23), 1889–1893.

Murphy E.A. (1981) *Skepsis, Dogma and Belief: Uses and Abuses in Medicine.* Johns Hopkins University Press, Baltimore.

Nelsen D.A., Hartley D.A., Christianson J., Moscovice I. & Chen M.M. (1994) The use of new technologies by rural family physicians. *Journal of Family Practice* **38** (5), 479–485.

Nelson H.S., Arenson J. & Reisman R. (1993) A prospective assessment of the remote practice of allergy: comparison of the diagnosis of allergic disease and the recommendations for allergen immunotherapy by board-certified allergists and a laboratory performing in vitro assays. *Journal of Allergy and Clinical Immunology* **92** (3), 380–386.

Norcini J.J., Kimball H.R. & Lipner R.S. (2000) Certification and specialization: do they matter in the outcome of acute myocardial infarction? *Academic Medicine* **75** (12), 1193–1198.

Norcini J.J., Lipner R.S., Benson J.A. Jr & Webster G.D. (1985) An analysis of the knowledge base of practicing internists as measured by the 1980 recertification examination. *Annals of Internal Medicine* **102** (3), 385–389.

Norcini J.J., Webster G.D., Grosso L.J., Blank L.L. & Benson J.A. (1987) Ratings of residents' clinical competence and performance on certification examination. *Journal of Medical Education* **62** (6), 457–462.

Pearce W.H., Parker M.A., Feinglass J., Ujiki M. & Manheim L.M. (1999) The importance of surgeon volume and training in outcomes for vascular surgical procedures. *Journal Vascular Surgery* **29** (5), 768–776.

Pearson K. (1911) *Grammar of Science*. p. 51. J. M. Dent & Sons, Ltd, London.

Ramsey P.G., Carline J.D., Inui T.S., Larson E.B., LoGerfo J.P. & Wenrich M.D. (1989) Predictive validity of certification by the American Board of Internal Medicine. *Annals of Internal Medicine* **110** (9), 719–726.

Sackett D.L. (1979) Bias in analytic research. *Journal of Chronic Diseases* **32** (1–2), 51–63.

Schwartz W.B. & Mendelson D.N. (1989) Physicians who have lost their malpractice insurance: their demographic characteristics and the surplus-lines companies that insure them. *JAMA* **262** (10), 1335–1341.

Sharp L.K., Bashook P.G., Lipsky M.S., Horowitz S.D. & Miller S.H. (2002) Specialty board certification and clinical outcomes: the missing link. *Academic Medicine* **77** (6), 534–542.

Simpson M.A. (1976) Medical student evaluation in the absence of examination. *British Journal of Medical Education* **10** (1), 22–26.

Sloan F.A., Mergenhagen P.M., Burfield W.B., Bovbjerg R.R. & Hassan M. (1989) Medical malpractice experience of physicians: Predictable or haphazard? *JAMA* **262** (23), 3291–3297.

Tussing A.D. & Wojtowycz M.A. (1993) The effect of physician characteristics on clinical behavior: Cesarean section in New York State. *Social Science and Medicine* **37** (10), 1251–1260.

Wood P.D.P. (1977) On the importance of correct orientation to sex in cattle. *Statistician* **26** (4), 304–306.