Learning in practice

Medical students' application of published evidence: randomised trial

Alan Schwartz, Jordan Hupert

Department of Medical Education (M/C 591), 808 South Wood Street, 986 CME, University of Illinois at Chicago, Chicago, IL 60612-7309, USA

Alan Schwartz *assistant professor*

Jordan Hupert *assistant professor*

Correspondence to: A Schwartz alansz@uic.edu

BMJ 2003;326:536–8

Abstract

Objectives To determine how medical students apply research evidence that varies in validity of methods and importance of results to a clinical decision.
Design Students examined a standardised patient with a whiplash injury, decided whether to order a cervical spine radiograph, and rated their confidence in their decision. They then read one of four randomly assigned variants of a structured abstract from a study of a decision rule that argued against such a procedure in this patient. Variants factorially combined two levels of validity of methods (prospective cohort or chart review) with two levels of importance of results (high sensitivity or high specificity rule). After reading the abstract, students repeated their choice and rated their confidence.
Setting Academic medical centre in the United States.
Participants 164 graduating medical students.
Main outcome measures Proportion of students in each group whose beliefs shifted or stayed the same.
Results When abstracts were of low importance students were more likely to shift their beliefs in favour of radiography, which was not supported by the evidence (odds ratio 3.42, 95% confidence interval 1.10 to 10.66). Neither methodological validity nor the interaction between validity and importance influenced decision shift. Few students acquired all necessary clinical data from the patient.
Conclusions Although the students could apply concepts of diagnostic testing, greater focus is needed on appraisal of validity and application of evidence to a particular patient.

Introduction

Evidence based medicine requires that clinicians learn to appraise findings in literature critically to determine their validity and importance. Although these skills are increasingly taught to medical students and residents,[1] a review of publications from 1973 to 1998 found that few reported sound evaluations of educational outcomes and none of actual or hypothetical clinical judgments.[2]

The impact of evidence can be inferred from the degree to which the evidence causes clinicians to change their belief in what they ought to do for a hypothetical patient.[3] We manipulated validity of methods and importance of results in a paper abstract presented to a group of graduating medical students after they had seen a standardised patient. We hypothesised that more important results should lead students to change their management beliefs, but only in so far as the results are derived from a study with high methodological validity.

Although over 75% of accredited US and Canadian medical schools use standardised patients or objective structured clinical examinations for teaching and assessing medical students,[4] in a literature search we found only one clear example of the use of standardised patients to evaluate evidence based medicine in clinical decisions. Bradley and Humpris found that students who followed evidence based recommendations received higher ratings from the standardised patients for quality of communication than students who did not.[5] A limitation was that students were not randomly assigned to the critical appraisal intervention.

Methods

Participants

All 164 fourth year graduating medical students at the University of Illinois at Chicago participated in the study. Participants examined the standardised patient as one part of an objective structured clinical examination consisting of 10 standardised patients. Exams were administered at the university's clinical performance centre. Although exam participation is required, exam results are used only formatively and have no impact on students' grades or progress.

Interventions

In the exam station, the patient was John, a 21 year old college student brought to the emergency department about 90 minutes after a 25 mph motor vehicle rear end collision. He had been a front seat passenger and was wearing a seatbelt. A few minutes before the ambulance arrived John had noticed a dull, constant pain in his neck. He was placed in a cervical collar, strapped to a spine board, and brought to the emergency department. This is the setting in which the student saw him.

John reported having dull, constant, generalised posteriolateral neck pain of mild intensity. He had no limb weakness, numbness, or paraesthesia. He remained in the cervical collar unless the student removed it. John's actual diagnosis was hyperextension-hyperflexion (whiplash) injury, and relevant differential diagnoses included soft tissue injury, cervical fracture, cervical dislocation, and instability of cervical ligaments.

BMJ: first published as 10.1136/bmj.326.7388.536 on 8 March 2003. Downloaded from http://www.bmj.com/ on 19 April 2024 by guest. Protected by copyright.

Students were directed to perform a focused physical examination and take the history. After the encounter, students were asked if they would request a radiograph of the cervical spine. Students responded either "yes" or "no" and rated their certainty in the correctness of their decision on a 16.8 cm visual analogue scale anchored by 50% (totally uncertain—flipping a coin would be as good as asking me) and 100% (totally certain that my decision is correct).

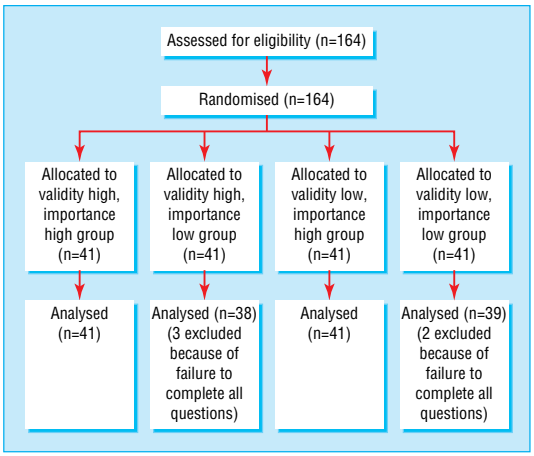
Students were then told that another student on their team had given them an abstract from a recent article in a peer reviewed journal about diagnostic criteria for cervical spine injuries requiring radiography. The abstract presented was based on a report of the Canadian spine rule for radiography in alert and stable trauma patients.[6] The abstract reported a prospective cohort study from which a three question decision rule was developed that has 100% sensitivity and 42.5% specificity for identifying clinically important cervical spine injuries. Under the decision rule, the standardised patient would not require radiography.

The authors, journal, and name of the rule were removed. Four versions of the abstract were then created based on a 2×2 (validity of methods × importance of results) factorial design. Levels of validity were high (the originally reported prospective cohort design) or low (a retrospective chart review). Levels of importance were high (the originally reported sensitivity and specificity) or low (42.5% sensitivity and 100% specificity, unsuitable for ruling out cervical spine injuries without radiography).

After reviewing the abstract, students repeated their decision and rating of certainty and wrote a brief statement about why their decision or certainty did or did not change. Students did not have the opportunity to re-examine the patient after reading the abstract.

Outcomes—Responses were scored by computing a certainty score for ordering a cervical radiograph both before and after exposure to the abstract. If the decision was "order radiograph" the certainty score was the certainty in the decision to order. If the decision was "do not order radiograph" the certainty score was 100% minus the certainty in the decision not to order. Accordingly, certainty scores ranged from 0 to 100. Responses were categorised into three patterns: no change (same certainty score before and after evidence), correct shift (certainty score decreased after evidence), or incorrect shift (certainty score increased after evidence). The standardised patient completed a checklist after the encounter. It included questions designed to determine whether the student had elicited information required to apply the decision rule: asking about tingling in the limbs, palpating the neck midline and muscles on each side after removing the cervical collar, and asking the patient to move his head from side to side.

Hypotheses—Our primary hypothesis was that the importance of the literature findings would predict changes in certainty score for ordering a radiograph—students would make fewer incorrect shifts when the rule had 100% sensitivity than when it had 100% specificity—but that this effect would be moderated by validity of methods. When such validity was low, we expected a less pronounced effect of importance of results than when it was high.



Participant flow diagram

Randomisation—We used block randomisation strategy to randomly assign students to receive one of the four versions of the abstract. A computer program generated 41 sets of the four versions, each randomly ordered. The sets were stacked and given to the testing centre, where each student received the next form from the top of the stack.

Statistical analysis—Multinomial logistic regression was used to predict response pattern (no change, incorrect shift, or correct shift) from result importance (high or low), validity of methods (high or low), and the interaction of validity and importance. The "correct shift" pattern was used as the baseline category for the regression.

Ethical issues—The UIC Institutional Review Board approved the study. Every student received feedback on the exam that included a debriefing page related to this patient and explained the study and abstract they received.

Results

All 164 students were recruited between 19 February 2002 and 3 April 2002. Data from 159 students were analysed for the primary outcome in an intention to treat fashion. The figure shows the allocation of participants.

On average, most students initially favoured ordering a radiograph (table 1). Groups did not differ significantly in their initial mean certainties. After the evidence the mean certainty score dropped significantly in all groups, both because students who favoured radiography became less certain and because students who did not favour radiography became more certain in their initial decision.

Most students correctly shifted their beliefs, particularly when they had read high importance abstracts (table 2). Multinomial logistic regression showed that when presented with the low importance (low sensitivity) abstract, students were more likely to incorrectly shift

Table 1 Mean (SD) certainty scores before and after students read evidence

	Before evidence		After evidence	
	Importance high (100% sensitivity)	Importance low (100% specificity)	Importance high (100% sensitivity)	Importance low (100% specificity)
Validity high (cohort)	82.2 (30.7)	80.4 (29.1)	60.0 (38.7)	55.3 (42.7)
Validity low (review)	83.8 (26.8)	82.7 (32.0)	54.9 (39.0)	63.3 (39.5)

BMJ: first published as 10.1136/bmj.326.7388.536 on 8 March 2003. Downloaded from http://www.bmj.com/ on 19 April 2024 by guest. Protected by copyright.

Table 2 Shifts in certainty scores after student read evidence

	Importance high (100% sensitivity)		Importance low (100% specificity)	
	Validity high (cohort) (n=41)	Validity low (review) (n=41)	Validity high (cohort) (n=38)	Validity low (review) (n=39)
Correct	31	34	20	27
No change	8	4	9	8
Incorrect	2	3	9	4

Table 3 Categorisation of open ended statements about why decision or certainty did or did not change. Figures are numbers (percentage) of students whose statement included each concept

Concept mentioned in statement*	No (%) mentioning
Clinical criteria from C-spine decision rule[6]	73 (46)
Sensitivity or specificity of rule:	
Overall	51 (32)
Used correctly	35
Used incorrectly	16
Nature of study design:	
Overall	8 (5)
Used correctly	8
Used incorrectly	0
Request for *x* ray examination to avoid potential lawsuits	7 (4)

*May include multiple concepts.

their beliefs in favour of radiography (odds ratio 3.42, 95% confidence interval 1.10 to 10.66). Neither validity of methods (1.46, 0.46 to 4.59) nor the interaction term (4.15, 0.43 to 40.27) was a significant predictor of incorrect shift, and no factor predicted the "no change" pattern. A multinomial ordinal logistic regression analysis produced the same pattern of results.

The patient checklist showed that 107 (67%) students asked about tingling in the limbs, 33 (21%) palpated the neck midline, and 27 (17%) asked the patient to rotate his head. In total, 17 (11%) students completed all three manoeuvres. Post hoc analyses introduced dummy variables representing completion of each manoeuvre into the regression. No significant effects of the completion of any manoeuvre on decision shift were found.

Exploratory analysis of the students' open ended statements was consistent with these findings (table 3).

Discussion

When a physician is exposed to a new piece of relevant evidence from the literature, uncertain beliefs ought to be influenced by the evidence, if the evidence is accepted as valid. Our medical students were influenced in their decision by exposure to clinically important evidence. The students who received the guideline with reported high sensitivity were significantly less likely to alter their beliefs erroneously than students who received the same guideline but with reported high specificity. However, the influence of the evidence did not depend on whether the guideline was produced from a prospective cohort study or a chart review.

Our findings should be interpreted with caution. The sample was limited to a single large medical school, albeit one in which evidence based medicine has received increasing time on the curriculum. Time constraints permitted the inclusion of only one case variant per student, and thus it was not possible to evaluate student specific variability. The difference in validity between a prospective and retrospective study in this case may not be sufficient to merit a difference in clinical judgment even when it is recognised.

Students did not re-examine the patient after they received the evidence. Although this is common in standardised patient assessments and in some clinical settings, in other clinical settings the evidence might guide a practitioner, particularly a student, to seek further information from the patient rather than commit to a decision about management.

Because assessment often drives student learning, integration of evidence based medicine into the curriculum requires integration into assessment, including standardised patient examinations. The results from this first application of the procedure are both reassuring and troubling. The students correctly applied the "SnNout" principle—a sensitive test with a negative result rules out the target condition.[7] However, they applied the decision rule without the clinical findings required to determine if the rule was applicable to their particular patient and paid no attention to the design of the research on which the evidence was based. These students may be learning to recognise salient results at the expense of considering how the results were obtained or whether the findings required to apply the results are present in the patient.

What is already known on this topic

Evidence based medicine is increasingly emphasised and taught in medical schools

Few studies have assessed the ability of physicians to apply literature findings to clinical decisions

What this paper adds

In making decisions about ordering investigations during a standardised patient exam students were sensitive to the importance of results

This effect was not moderated by validity of the study that produced the results or whether the students had collected enough information to apply the results

We thank Rachel Yudkowsky, director, Clinical Performance Center; Barbara Eulenberg, standardised patient coordinator; and Robert Kiser and Thad Anzur, who portrayed the standardised patients for the study. Georges Bordage and Julie Goldberg provided helpful comments on a draft of the manuscript.

Contributors: AS and JH developed the study and wrote and revised the paper. JH developed the standardised patient cases. AS carried out the analysis and is guarantor.

Funding: None.

Competing interests: None declared.

1 Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ* 1998;158:177-81.
2 Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med* 1999;74:686-94.
3 Schwartz A, Hupert J, Hasnain M, Elstein AS, Noronha P, Gaebeler C, et al. Evaluating evidence-based decision making skills: development and validation of an instrument [abstract No 408]. *Pediatr Res* 2001;49(suppl):74A.
4 Association of Academic Medical Colleges. *LCME annual medical school questionnaire, 1999-2000.* www.aamc.org/data/gq/allschoolsreports/2000.pdf (accessed 10 Dec 2002).
5 Bradley P, Humpris G. Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. *Med Educ* 1999;33:815-7.
6 Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA* 2001;286:1841-8.
7 Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence based medicine: how to practice and teach EBM.* 2nd ed. Edinburgh: Churchill Livingstone, 2000.

(Accepted 21 November 2002)

BMJ: first published as 10.1136/bmj.326.7388.536 on 8 March 2003. Downloaded from http://www.bmj.com/ on 19 April 2024 by guest. Protected by copyright.