Is a Journal Club Effective for Teaching Critical Appraisal Skills?

A Controlled Trial With Residents in Psychiatry

Cynthia H. Y. Fu, M.D., M.Sc.
Brian Hodges, M.D., M.Ed.
Glenn Regehr, Ph.D.
David S. Goldbloom, M.D.
Paul Garfinkel, M.D.

This study examined the effectiveness of a journal club for teaching critical appraisal skills to residents in psychiatry and their transfer of those skills to clinical scenarios. Twelve residents participated in a 12-week journal club, and 12 residents were matched control subjects. Following the journal club, there remained no difference in performance between the two groups, although two-thirds of the journal club residents did show an improvement or no change, compared with one-third of the control residents. Furthermore, there was a trend for the journal club residents to require less time to review an article. This format of a journal club may not *be an effective method for teaching critical appraisal skills to residents in psychiatry. (Academic Psychiatry 1999; 23: 205–209)*

A key element of continuing medical education is the ability to critically appraise literature and to apply this knowledge in clinical practice. In its report entitled "Physicians for the Twenty-First Century," the Panel on the General Professional Education of the Physician has targeted the identification and critical appraisal of relevant literature as vital to the training of competent physicians (1). Journal clubs are a time-honored tradition for learning the latest in medical research and teaching of critical appraisal skills. Yet despite the tradition of journal clubs, empirical evidence for their effectiveness is sparse.

Studies have examined teaching critical appraisal skills to medical students and residents in medicine, family medicine, and surgery, using controlled trial designs of two groups with random or near random assignment (2–9). Measurements varied among the studies, but most were composed of self-report satisfaction questionnaires and written multiple-choice questions or short essays, consisting of 10 to 20 items (2–9). All measurements were made before and immediately after the trial, with no further follow-up. In a recent review, Norman and Shannon reported that gains in knowledge were consistent at the medical student level (mean ± SD, gain: 17.0 ± 4.0% on written knowledge tests), but were small at the residency level (gain: 1.3 ± 1.7%) (10). However, no study has attempted to examine an application to clinical practice. Furthermore, although the most significant

Dr. Fu was the Chief Resident, Clarke Site, Centre for Addiction and Mental Health (CAMH) and is currently a Clinical Research Fellow, Institute of Psychiatry and Maudsley Hospital, London UK. Dr. Hodges is the Vice-Chair of Education and Assistant Professor, Department of Psychiatry, University of Toronto; and Dr. Regehr is the Associate Director, Centre for Research in Education, University Health Network, and Associate Professor, Faculty of Medicine, University of Toronto. Dr. Goldbloom is the Physician-in-Chief, CAMH, and Professor of Psychiatry; and Dr. Garfinkel is the President and CEO, CAMH, and Professor of Psychiatry, Department of Psychiatry, University of Toronto, Canada. Address correspondence and reprint requests to Dr. Fu, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, United Kingdom; e-mail: c.fu@iop.kcl.ac.uk.

effects have been demonstrated with medical students, critical appraisal skills are more relevant during clinical practice.

Thus, we have attempted to examine the teaching of critical appraisal skills at the residency level in psychiatry, a group that has not been previously studied and for whom such skills may be particularly important given the growth of biological and therapeutic advances. We have further attempted to study the extent to which such skills may then be applied in day-to-day patient care. We hypothesized that 1) critical appraisal skills could be taught to psychiatry residents; 2) a 12-week journal club would be sufficient; and 3) these skills would be demonstrated in clinical case scenarios and, thus, possibly in patient care. The measurement involving clinically relevant scenarios was novel and added a validity that had been lacking in prior studies.

METHODS

Residents (postgraduate year [PGY]-1 to PGY-5) at University of Toronto Department of Psychiatry participated, with approval given by the University of Toronto Ethics Review Committee. The experimental group consisted of 12 residents with a mean ± SD PGY-2.6 ± 0.8, and the matched control group was 12 residents with a mean ± SD PGY-3.1 ± 1.2, with no significant difference between the groups ($t_{22} = 1.23$, NS). The experimental trial consisted of a weekly journal club, each seminar lasting 1.5 hours, for 12 weeks. Baseline measurements were collected for journal club residents in September 1996 and in a staggered time schedule, from October to November 1996, for control residents who were recruited from two hospitals in order to match the journal club residents from one hospital. Follow-up measurements were made at the end of the course for the journal club residents and at a 12-week follow-up for the control residents. Data for journal club residents were included if they attended a minimum of half the sessions.

A professor of psychiatry led the journal club, and for each seminar, one resident was required to select a clinically relevant article, distribute it the week prior, and critically review it during the seminar with respect to its introduction, study design, biases, sample size, generalizability, statistical methods, results, discussion, conclusions, and other areas relevant to the article. All residents were encouraged to participate freely while the professor would facilitate and elaborate on pertinent aspects. Four of the sessions included 45-minute presentations on basic statistical concepts by an expert staff psychologist (G.R.). The control group did not participate in a similar journal club during this period.

The measurements consisted of the following:

1. A questionnaire designed specifically for the study that surveyed current practices and opinions (Table 1). Some measurements had anchored 5-point rating scales, with responses of agreement or importance being "1" = "not at all," "3" = "some," and "5" = "very much." Others, for example "How many hours a week do you spend reading journal articles?" consisted of endorsing responses ranging from "<1 hour" to ">2 hours," and of dichotomous responses, such as the impact of the journal club.

2. A 19-item multiple-choice questionnaire (MCQ) to evaluate the resident's ability to critically assess a four-page article, a treatment, and outcome study selected from the literature. Questions included examination of the reputation of the authors, introduction, method, data analysis, results, discussion, conclusion, clinical applicability, and an overall evaluation.

The answers to the MCQ's were determined by the "Delphi" method, which entailed distribution of the article to six staff psychiatrists who made specific comments on each aspect. The MCQ was designed from a consensus of their comments, with confirmation of the answers.

At follow-up, a different article was chosen, but with a similar experimental design, and MCQ items were determined by using the same format. Both tests also included a four-item MCQ of basic statistical concepts.

3. Seven clinical case scenarios were developed for the purpose of evaluating knowledge of treatment decisions, prognosis of illness, basic pathophysiology, and epidemiology. An example: "After a single episode of mania in a male in his early 20s, with no family history of bipolar disorder, how long do you maintain prophylactic mood-stabilizing medication?" Answers were selected from five possible responses.

4. In the case scenarios, the residents also indicated the source of their knowledge, with options of staff psychiatrists, residents, journals, textbooks, pharmaceutical companies, newspapers, and not ap-

plicable for the questions for which they did not know the answer. The residents were permitted to make more than one selection.

The same case scenarios were given at baseline and at the 12-week follow-up for the purpose of examining for a change in knowledge and, particularly, the source of the residents' knowledge. Although the residents were aware of the purpose of the study, they were not aware of the purpose of this additional component.

STATISTICAL ANALYSIS

Between-group comparisons of mean year of residency, mean amount of time required to review the articles, percentage score of correct answers of critical appraisal skills and knowledge of statistics, and percentage score of case scenario responses were made by independent samples *t*-tests. Comparisons of the median amount of time required to review the articles between the groups were made by Mann-Whitney *U* tests. Within-group comparisons of baseline and follow-up responses were made by paired *t*-tests. Correlations of follow-up performance with motivation score on the questionnaire and journal club attendance were assessed by Pearson's two-tailed correlation analysis.

A power calculation based on an estimate of 15 residents in the experimental group and a matched control group, as prior studies had group sizes from 12 to 51 residents, with an α error of 0.05 and β error of 0.20, indicated an effect size of 0.74 (6–9). Given an acceptable within-group test-retest reliability, an effect size of 0.74 would be a reasonable expectation if the experimental condition were to have a clinically significant effect.

RESULTS

At baseline, both groups reported that critical appraisal skills were "important" and that their own abilities were "adequate" (Table 1). The journal club residents reported that they were "motivated" to participate (Table 1). However, they attended a mean $\pm$ SD of 8.3 ± 1.9 (range: 6–12) of the 12 sessions, citing clinical workload and vacation time; yet the journal club residents believed that the duration of the course was sufficient, rating the duration as 2.7 ± 1.0 on a five-point anchored rating scale, with "3" representing "just long enough." The control residents confirmed that they did not have any formal teaching of critical appraisal skills during the 12-week follow-up.

At baseline, the journal club residents required a mean of 13.8 ± 5.6 minutes, median 13 minutes, to review the article, and the control residents required a mean of 13.7 ± 10.3 minutes, median 13 minutes, with no significant difference between the groups ($t_{22} = 0.05$, NS). At follow-up, the journal club residents required a mean of 14.1 ± 6.1 minutes, median 10 minutes, whereas the control residents required mean of 18.0 ± 6.2 minutes, median 20 minutes. The mean times at follow-up showed no significant difference between the groups ($t_{19} = 1.45$, NS), although median times showed a trend toward significance (Mann-Whitney $U = 31$, $P = 0.11$) (Table 2). At baseline and at follow-up, there were no significant differences between the groups in their abilities, neither in critical appraisal skills and knowledge of statistics nor in clinical case scenarios (Table 2). In the journal club group, 7 of the 12 residents showed an improvement or no change in their critical appraisal skills and statistics knowledge, compared with 4 of the 12 residents in the control group. Correlation of their

TABLE 1. Baseline questionnaire using a 5-point scale with level of ability, importance, or skill as "1" = "not at all," "3" = "some," and "5" = " very much," with no significant difference between the groups on any measurement

Baseline questionnaire	Journal Club Residents	Control Residents
Median time spent on weekly review of literature (range)	2 hours (⟨1 to⟩2 hours)	1 hour (⟨1 to⟩2 hours)
Perceived importance of critical appraisal skills (mean ± SD)	4.3 ± 1.1	4.2 ± 0.8
Rating of resident's own critical appraisal skills (mean ± SD)	3.1 ± 0.5	3.2 ± 0.8
Rating of supervisors' use of literature (mean ± SD)	3.2 ± 1.1	3.5 ± 1.1
Rating of other residents' use of literature (mean ± SD)	2.8 ± 0.4	2.5 ± 0.5
Motivation to participate in course (mean ± SD)	3.7 ± 1.0	3.1 ± 0.5

follow-up performance to their level of motivation was $r = -0.49$ (NS, df = 9) for the experimental group and $r = -0.25$ (NS, df = 9) for the control group, with no significant correlation for either group, although data were absent, as some residents did not fully complete this questionnaire. There was no significant correlation with number of sessions attended and follow-up performance for the journal club residents ($r = 0.36$, NS, df = 11).

Fifty-eight percent of the journal club residents stated that the course had a positive impact on their use of the literature, and 42% stated that it had an impact on their work. At baseline, the majority of the residents chose staff psychiatrists as their main source of knowledge. Interestingly, at follow-up, both groups indicated a reduction in a primary reliance on staff psychiatrists (Table 2). However, the journal club residents did not report a change in their use of journals, nor of textbooks. In contrast, the control residents indicated an increased use of journals and a decreased use of textbooks. At both measurements, neither group selected pharmaceutical companies, newspapers, or other residents as a source of knowledge.

DISCUSSION

We have examined the time-honored tradition of a journal club for teaching critical appraisal skills to residents in psychiatry and have attempted to further examine its application by using clinical case scenarios. At baseline, both the journal club and control groups were equivalent in their abilities to critically appraise an article, knowledge of statistics, and performance on clinical case scenarios. Despite their attendance and perception of the journal club as having a positive impact, at follow-up the journal club residents did not perform any better than the control residents, although a greater number of the former did demonstrate an improvement or no change relative to the control residents. Interestingly, the control residents seemed to require more time to review and appraise the follow-up article than the journal club residents. Over half the journal club residents reported that the course had an impact on their use of the literature, whereas 42% felt there was an impact on their work. However, it was the control residents who cited more journal references at follow-up; their journal use increased and exceeded that of the journal club residents. These findings were contradictory to our initial hypotheses.

Limitations of this study include the small sample size, although it is comparable to prior studies. This study used a controlled trial design, with the journal club and control residents divided by hospital sites. Although possible variability in hospital resources, including staff psychiatrist accessibility, may be a confounding variable, this effect was accounted for by an absence of baseline performance differences between the groups who had been working at their respective sites for several months. As well, the residents and chief residents of each hospital indicated that facilities were comparable, and the residents had full access to all hospitals. Furthermore, the residents

TABLE 2. Performance on tests of critical appraisal skills and statistics and on clinical case scenarios at baseline and at a 12-week follow-up

	Journal Club Residents		Control Residents	
Measurement	Baseline $N = 12$	Follow-up $N = 12$	Baseline $N = 12$	Follow-up $N = 10$
Mean ± SD time for article (minutes)	13.8 ± 5.6	14.1 ± 6.1	13.7 ± 10.4	18.0 ± 6.2
Median time for article (minutes)	13	13	10	20
Critical appraisal and statistics (percentage correct: mean ± SD)	66.7 ± 14.5	64.4 ± 10.4	63.6 ± 11.1	67.4 ± 12.7
Clinical scenarios (perentage correct: mean ± SD)	48.2 ± 15.2	48.1 ± 27.3	42.9 ± 19.3	50.8 ± 26.8
Source of knowledge (%):[a]				
Staff psychiatrists	54	49	54	44
Journals	21	23	14	21
Textbooks	13	12	29	16
Pharmaceutical-sponsored	0	0	0	0
Other residents	0	0	0	0
No source indicated	13	18	26	25

Note: N = number of residents; [a]Percentages may add up in excess of 100% as more than one source could have been identified.

at each PGY had similar clinical duties, and all residents shared equally in call duties, which were distributed across hospitals. Another limitation is that we collected 12-week follow-up data; long-term measurements were not made, although their usefulness is questionable, as the immediate follow-up showed no group differences. A possible bias in the study may be that the control residents were aware of the journal club at another hospital site. Although the University of Toronto is one of the largest psychiatry residency programs in North America, residents have significant interaction through shared call duties and seminars among sites, and we hypothesized that residents would discuss the study. An alternative explanation would have been a placebo-arm of weekly 1.5-hour general meetings, which residents would likely have stopped attending. We believed the best compromise was to be upfront about the study. Nonetheless, the journal club residents in our study did not perform better than the control residents.

The results suggest that the journal club, as shown in this study, is *not* an effective method for teaching critical appraisal skills to residents in psychiatry. Our findings add to the existing literature that has shown that gains in critical appraisal skills are minimal to none during residency, irrespective of the duration of the formal teaching (10). In addition, the journal club residents did not describe an increased use of the literature, which corresponds to an earlier study finding that internal medicine residents read fewer articles following a journal club course (7).

The format of a journal club is highly regarded as a way to teach critical appraisal skills. However, contrary to our hypothesis, this study corroborates the literature that journal clubs are *not* effective at the residency level. Perhaps residents have already acquired some skills, and further teaching produces diminishing returns. As well, the effectiveness of journal clubs may lie in other facets, such as improving collegiality and communicating new literature. In further studies, double-blind placebo-control trials are not practical, but perhaps a "wait-list" group, which eventually has journal club participation, may be feasible. Finally, we suggest that the learning of critical appraisal skills may be promoted by the degree of personal importance to residents, and this factor is often in a form that affects their evaluations (11).

This paper was presented at the Ontario Psychiatric Association Annual Meeting, Toronto, Ontario, Canada, January 1997.

The work was done at the Clarke Site, Centre for Addiction and Mental Health, 250 College Street, Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada.

Dr. Fu gratefully acknowledges the support by an Educating Future Physicians of Ontario Fellowship.

References

1. Physicians for the Twenty-First Century. The GPEP Report: Report of the Panel on the General Professional Education of the Physician and College Preparation for Medicine. Washington, DC, Association of American Medical Colleges, 1984
2. Heller RF, Peach H: Evaluation of a new course to teach the principles and clinical applications of epidemiology to medical students. Int J Epidemiol 1984; 13:533–537
3. Radack KL, Valanis B: Teaching critical appraisal and application of medical literature to clinical problem solving. Journal of Medical Education 1986; 61:329–331
4. Bennett KJ, Sackett DL, Haynes RB, et al: A controlled trial of teaching critical appraisal of the clinical literature to medical students. JAMA 1987; 257:2451–2454
5. Frasca MA, Dorsch JL, Aldag JC, et al: A multidisciplinary approach to information management and critical appraisal instruction: a controlled study. Bull Med Libr Assoc 1992; 80:23–27
6. Gehlbach SH, Bobula JA, Dickinson JC: Teaching residents to read the literature. Journal of Medical Education 1980; 55:362–365
7. Linzer ML, Brown JT, Frazier LM, et al: Impact of a medical journal club on house staff reading habits, knowledge and critical appraisal: a randomized controlled trial. JAMA 1988; 260:2537–2541
8. Kitchens JM, Pfeifer MP: Teaching residents to read the medical literature: a controlled trial of a curriculum in critical appraisal/clinical epidemiology. J Gen Intern Med 1989; 4:385–387
9. Seelig CB: Affecting residents' literature reading attitudes, behaviors and knowledge through a journal club intervention. J Gen Intern Med 1991; 6:330–334
10. Norman GR, Shannon SI: Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. Canadian Medical Association Journal 1998; 158:177–181
11. Swanson DB, Norman GR, Linn RL: Performance-based assessment: lessons from the health professions. Education Research 1995; 24:5–11