

Research Paper ■

A Comparison of the Effects of Computer and Manual Reminders on Compliance with a Mental Health Clinical Practice Guideline

DALE S. CANNON, PHD, STEVEN N. ALLEN, PHD

Abstract **Objective:** To evaluate the relative effectiveness of computer and manual reminder systems on the implementation of a clinical practice guideline.

Design: Seventy-eight outpatients in a mental health clinic were randomly assigned within clinician to one of the two reminder systems. The computer system, called CaseWalker, reminded clinicians when guideline-recommended screening for mood disorder was due, ensured the fidelity of the diagnosis of major depressive disorder to criteria of the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV), and generated a progress note. The manual system was a checklist inserted in the paper medical record.

Measures: Screening rates for mood disorder and the completeness of the documentation of which DSM-IV criteria were met by patients who were said to have major depressive disorder were compared.

Results: The CaseWalker, compared with the paper checklist, resulted in a higher screening rate for mood disorder (86.5 vs. 61 percent, $P = 0.008$) and a higher rate of complete documentation of DSM-IV criteria (100 vs. 5.6 percent, $P < 0.001$).

Conclusions: In an outpatient mental health clinic, computer reminders were shown to be superior to manual reminders in improving adherence to a clinical practice guideline for depression.

■ JAMIA. 2000;7:196–203.

Many clinical practice guidelines have been promulgated over the past decade. Among the factors driving the increased development of clinical practice guidelines is the belief that they can both improve the quality of care and reduce the cost of care.^{1,2} If guidelines are to accomplish their intended purposes, health care providers must follow them. A review of the English literature between 1981 and 1990 on guideline compliance found that the mean compliance rate with guideline recommendations was only 55 percent.³

Thus, there is room for improvement in guideline compliance.

The usual first step in implementing a clinical practice guideline is to educate practitioners about the guideline through some passive pedagogic technique such as lectures or publication of the guideline. Although such techniques may increase awareness of the guideline and may be necessary, they have not proved to be sufficient to change actual practice patterns.⁴

In this study, the effects of computer and manual reminders on compliance with a clinical practice guideline for depression were compared in an outpatient mental health clinic. Manual reminders currently are more widely used than computer reminders because they are so simple to implement, but computer reminders may be more effective. As noted later, the empirical evidence on the relative effectiveness of the

Affiliation of the authors: University of Utah, Salt Lake City, Utah.

Correspondence and reprints: Dale Cannon, PhD, Psychiatry Service (116OP), VAMC, 500 Foothill Boulevard, Salt Lake City, UT 84148; e-mail: <dale.cannon@med.va.gov>.

Received for publication: 6/25/99; accepted for publication: 10/15/99.

two methods is inconclusive. Computer reminder systems have been studied extensively in medical ambulatory care clinics, but we could find no published studies of their use in mental health settings.

Background

Two excellent systematic review articles^{5,6} and two meta-analyses^{7,8} based on randomized controlled trials of computer reminder systems conclude that computer reminders enhance physician adherence to clinical practice guidelines. An alternative to computer reminder systems is manual systems in the form of paper checklists, structured encounter forms, and flow sheets.⁹⁻¹² Although manual reminder systems have not been evaluated as thoroughly as computer systems, there is some evidence that manual systems increase preventive health screening by physicians, in comparison with screening rates obtained with no reminder system.¹³⁻¹⁷ Given that both computer and manual reminder systems are widely used and have some empirical support, a direct comparison of the two is indicated to determine whether the benefits of computer systems justify the time and effort required to implement them. The meta-analysis by Shea et al.⁸ considered four intervention conditions—computer reminders alone, manual reminders alone, both types of reminders combined, and no reminders. Each type of reminder system alone resulted in higher compliance with preventive medicine recommendations than did the no-reminder control condition. However, there was no significant difference between computer reminders alone and manual reminders alone. Contrary to the findings of Shea et al.,⁸ two randomized controlled trials of computer and manual reminder systems report significant benefits of computer reminders compared with manual reminders.^{18,19} The inconclusive evidence concerning the relative effectiveness of computer and manual reminders motivates further comparison of the two methods.

The present study was a comparison of the effects of computer and manual reminders on compliance with two requirements of a clinical practice guideline for major depressive disorder (MDD) that was developed by the Veterans Health Administration (VHA).²⁰ Two elements of the MDD clinical practice guideline were mandated in VHA. First, all patients seen at least three times in outpatient clinics within one year were to be screened for mood disorder. Second, the documentation of MDD diagnoses given to patients who were followed in outpatient mental health clinics was to specify the criteria for MDD, set forth in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition²¹ (DSM-IV), that the patient met.

Methods

Research Participants

The four research participants were senior clinicians employed by the Posttraumatic Stress Disorder (PTSD) Clinical Team at the Salt Lake Veterans Affairs Medical Center. One was a clinical psychologist, one was a registered nurse, one was a social worker, and one was an addiction therapist. They have an average of 16 years of professional experience and have worked with the PTSD Clinical Team for an average of two years.

Study Site

The PTSD Clinical Team, an outpatient clinic for patients with PTSD, was chosen as the study site for both design and practical reasons. The design consideration was a high incidence of MDD in patients with PTSD,²² which increased the opportunity to compare the two experimental methods in their effectiveness in documenting the DSM-IV diagnostic criteria for MDD that these patients met.

The PTSD Clinical Team was chosen as the study site for several practical reasons. First was the willingness of the PTSD Clinical Team staff to participate in the study, especially the willingness of the Coordinator of the PTSD Clinical Team to oversee the assignment of cases to experimental conditions. Second, the PTSD Clinical Team has permanent staffing, so the logistical and statistical problems associated with the rotation of residents and interns were avoided. Third was the location of the clinic, which was within a few feet of the offices of the investigator and his staff. The fourth reason was the availability of hardware and a local area network to support the computer arm of the study.

Measures

Two variables were used to assess the relative effectiveness of the two reminder methods. The first measure was the proportion of cases screened for mood disorder. The second measure was the proportion of cases for which the diagnosis of MDD was fully documented according to DSM-IV diagnostic criteria. For cases in the computer reminder condition, both measures were determined by checking the computer reminder database. In the paper condition, both measures were determined from the paper checklist.

Computer Reminder System

The computer reminder system was dubbed the CaseWalker, after the recommendation of the guide-

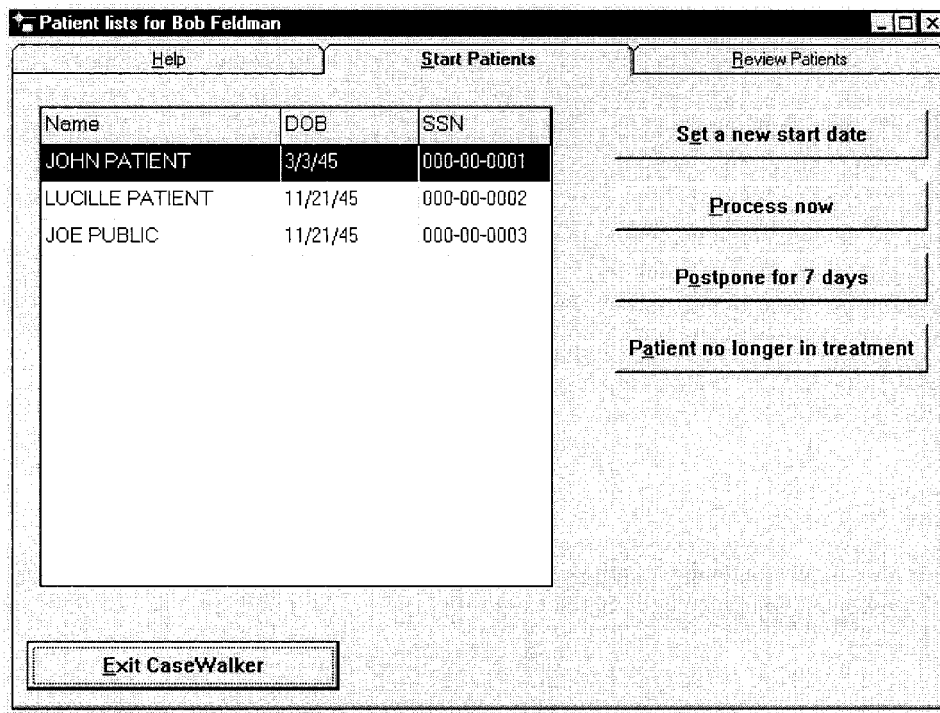


Figure 1 Reminder for patients requiring initial screening and assessment.

line developers that staff learn the logic of the guideline by “walking” cases through it. The CaseWalker generated reminders to screen patients for mood disorder, presented and scored the DSM-IV criteria for MDD, and created a progress notes based on answers given to questions derived from the guideline.

The CaseWalker platform was a Windows NT 4.0 local area network. The graphical user interface was written in Delphi 3. An Interbase database resided on the local network server. Expert C++, an inference engine and knowledge base, was used to encode the guideline algorithm.

Each clinician saw patients in his or her office, and each had a PC on his or her desk that ran the CaseWalker as well as the VHA electronic medical record (EMR). Use of the EMR increased over the course of the study. In the beginning, clinicians used it most often to view appointment schedules, and sometimes to view laboratory results. By the end of the study, they were using it for progress notes, laboratory orders, and medication orders. Thus, the clinicians were accustomed to using their computer in their clinical practice. However, it was not possible technically at the beginning of the study to integrate the EMR and the CaseWalker directly, so progress notes generated by CaseWalker had to be “cut and pasted” into the EMR.

Daily, the CaseWalker presented each clinician with a list of patients on the clinician’s caseload who needed

to be screened for mood disorder (Figure 1). For each patient on the reminder list, the user had the option of selecting a new reminder date, terminating the reminders, or processing the guideline. Users who opted to process the guideline for a patient were asked whether a four-item screening test for mood disorder was positive. An optional pop-up form (Figure 2) was available to administer and score the screening test. Alternatively, the clinician could simply enter the result (positive or negative) of the screening test.

If the mood disorder screening test was positive, an MDD diagnostic criteria checklist popped up automatically (Figure 3). These criteria include the criteria for a diagnosis of major depressive episode (e.g., depressed mood, insomnia, weight change, fatigue) as well as the rule-out criteria for MDD (e.g., no psychotic or schizo-affective disorder, manic or mixed episodes, substance-induced mood disorder, or normal bereavement). In addition, the diagnosis of MDD requires a determination that the depressive symptoms caused clinically significant distress or impairment. The user was required to use the checklist in making the diagnosis of MDD, and the diagnostic criteria were scored automatically by the Delphi program.

At the conclusion of a CaseWalker session, the clinician was given a chance to review the progress note that had been generated during the session. Each user entry during the session concatenated appropriate

Figure 2 Checklist for initial screening for mood disorder.

Assessment Tools

Brief Screen for Depression

In the past year, have you had two weeks or more during which you felt sad, blue or depressed; or when you lost all interest or pleasure in things that you usually cared about or enjoyed?

Have you had two years or more in your life when you felt depressed or sad most days even if you felt okay sometimes?

Have you felt depressed or sad much of the time in the past year?

How much of the time during the past week did you feel depressed?

Less than one day

1-2 days

3-4 days

5-7 days

OK Cancel

For Age < 60 yrs / For Age > 60 yrs / DSM-IV Criteria for MDE / DSM-IV Criteria for MDD / PTSD Criteria

Figure 3 First section of the checklist of DSM-IV criteria for major depressive disorder.

MDD CaseWalker

File Module

Status Report: PTSD Assessment | PTSD Outpatient | PTSD Inpatient

Patient referred with positive results on brief screen for both PTSD and Mood Disorder symptom (SU/A.)

Positive results

Page 1

Depression Assessment

DSM-IV Diagnostic criteria for MDE

A. At least five of the following symptoms have been present during the same 2-week period, nearly every day, and represent a change from previous functioning. At least one symptom must be either depressed mood or loss of interest or pleasure.

depressed mood (or alternatively can be irritable mood in children or adolescents)

Markedly diminished interest or pleasure in all, or almost all, activities

significant weight loss or weight gain when not dieting

insomnia or hypersomnia

psychomotor retardation or agitation

fatigue or loss of energy

feelings of worthlessness or excessive or inappropriate guilt

diminished ability to think or concentrate

OK Cancel

Does the patient for MDD?

Yes No

Flowchar Guideline Write Pro Determin

continued

For Age < 60 yrs / For Age > 60 yrs / DSM-IV Criteria / Primary Care Treatment / PTSD Crit

Begin Restart Load Ke

text to the progress note, which could be edited by the clinician. Then, the note could be either printed and filed in the paper medical record or pasted to an electronic progress note in the hospital information system.

Paper Checklist

The clinic clerk inserted the checklist into the assessment section of the paper medical record of each new patient assigned to the checklist arm of the study. Prior to each appointment, the paper medical record was to be pulled by clerical staff and made available to the clinician.

The paper checklist was three pages long. The first page contained the four-item mood disorder screening test. The next two pages contained the DSM-IV criteria for MDD in checklist form. If the screening test was positive, the clinician was to check all DSM-IV criteria that applied and determine whether the patient satisfied diagnostic criteria for MDD. These diagnostic criteria, of course, were exactly the same as those presented by the CaseWalker, and they were presented in the same order.

Procedure

Consecutive admissions to the caseloads of the four participating clinicians between Jan 5, 1998, and Oct 7, 1998, were assigned randomly to one of the two experimental conditions. The coordinator of the PTSD Clinical Team assigned 108 patients newly referred to the clinic to one of the four PTSD Clinical Team clinicians on a nonrandom basis. The most common consideration in assigning a new patient was the current caseload of each clinician, but sometimes patient characteristics determined clinician assignment. For example, a female patient with PTSD secondary to sexual trauma would be assigned to the female therapist for clinical reasons. Even though the assignment of cases across clinicians was not random, potential bias of the comparison of the two reminder methods was controlled for by the random assignment of cases to experimental condition within clinician. Randomization was based on a table of random numbers.

Because there was not always time to screen for mood disorder during the first session with a patient, only those cases seen by the assigned clinician at least twice were included in the data analyses. A total of 83 cases met the two-visit-minimum criterion, but five cases that were supposed to be in the CaseWalker group were dropped from the study because of a clerical error that resulted in the CaseWalker procedure never being initiated for these cases. A total of 37 and 41 patients were assigned to the CaseWalker and pa-

per checklist conditions, respectively. A chi-squared (χ^2) analysis indicated no significant difference in the number of cases per clinician and experimental condition in the sample of 78 patients who had at least two visits ($\chi^2 = 0.56, P = 0.91$).

Data Analyses

Analysis of variance (ANOVA)²³ was used to analyze the results. Independent variables were experimental condition (i.e., CaseWalker vs. paper checklist), clinician, and the experimental condition-by-clinician interaction. Dependent variables were the absence or presence of the indicated measure, scored as 0 (absent) or 1 (present). The effect size of significant group differences was determined by means of the *f* statistic.^{24,25} Cohen²⁴ designates an *f* value of 0.10 as small, 0.25 as medium, and 0.40 as large.

Results

More cases were screened for mood disorder in the CaseWalker condition (86.5 percent) than in the paper checklist condition (61 percent) (Figure 4). An ANOVA in which the occurrence or nonoccurrence of screening was the dependent variable was significant for both the group ($F(1,70) = 7.45; P = 0.008$) and clinician effects ($F(3,70) = 2.85; P = 0.044$). The *f* statistic for the group effect was 0.30, which indicates a medium effect size.²⁴ The group-by-clinician interaction was not significant.

The percentage of patients in each arm of the study who were said to have MDD and the percentage for whom DSM-IV criteria were fully documented are shown in Figure 5. The clinicians identified 46 percent of patients in the CaseWalker condition and 44 percent of those in the paper checklist condition as hav-

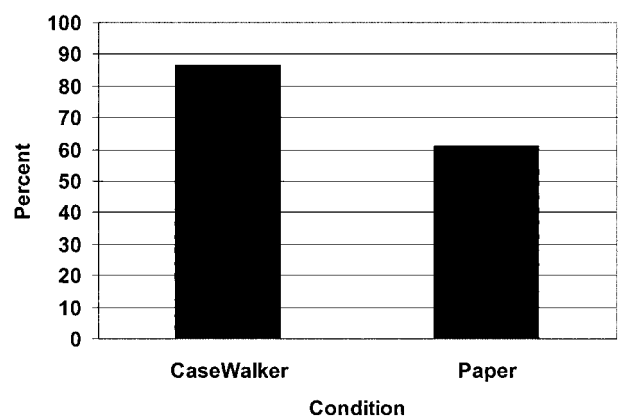


Figure 4 Percentage of cases screened for mood disorder in the CaseWalker and paper conditions.

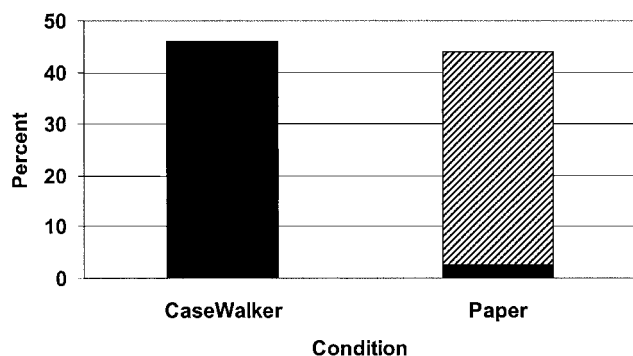


Figure 5 Percentage of cases in which major depressive disorder criteria were documented in the CaseWalker and paper conditions. Black indicates complete documentation; hatching, incomplete documentation.

ing MDD. For all 17 (100 percent) of the CaseWalker patients who were said to have MDD, the DSM-IV criteria that were met were fully documented. In the paper checklist condition, the criteria were fully documented for only 1 of 18 patients (5.6 percent) who were said to have MDD. An ANOVA across the 35 patients said to have MDD, in which the dependent variable was whether the diagnosis of MDD was fully documented, was significant for the group effect ($F(1,27) = 170.9; P < 0.001$). The f statistic for the group effect was 2.83, which is substantially higher than the 0.40 that Cohen²⁴ designates a large effect size.

Discussion

The results of this study clearly demonstrate that computer reminders were more effective than manual reminders in supporting clinician compliance with the VHA clinical practice guideline for MDD. This finding contrasts with the report by Shea et al.⁸ of no significant difference between computer and manual reminder systems. The failure of Shea et al. to find a significant effect, of course, is not the same thing as demonstrating that the effect does not exist. There are now three independent randomized controlled trials,^{18,19} in quite different clinical settings, that have found an advantage for computer reminder systems. Although further research would be required to specify the conditions under which the two types of reminders do and do not differ, they are likely to differ across a range of clinic settings and types of practice guidelines.

The present study does not demonstrate that manual reminders are ineffective, simply that they are less effective than computer reminders. The experimental design did not include a no-reminder control condition, which would have been necessary to assess the effectiveness of the manual system. It was not possible

to have a no-reminder control condition because of the VHA mandate that the MDD guideline be implemented and the fact that a paper checklist was the accepted standard means of complying with the mandate.

From a statistical perspective, the effect of the computer reminder, compared with the manual reminder, on screening for mood disorder was of medium magnitude.²⁴ The effect size for complete documentation of DSM-IV criteria for MDD was of large magnitude.²⁴ These medium and large effect sizes were observed in spite of an experimental design that may have attenuated group differences. Hunt et al.⁶ point out that a design in which individual clinicians are observed under multiple experimental conditions may underestimate experimental effects because the clinician's experience with one arm of the study may generalize to the other. For example, in the present study, if the CaseWalker made clinicians more aware of the requirement to screen patients for mood disorder and if that awareness increased their use of the checklist, then any differential effect the CaseWalker may have had on mood screening would have been attenuated. A study of computer-generated physician reminders provides evidence that such interactions occur between experimental conditions when clinicians participate in multiple arms of a computer reminder study.²⁶ The significant difference in the present study between the CaseWalker and the paper checklist, in spite of the conservative bias of the design, increases confidence that it is a robust finding.

In the paper checklist condition, the diagnosis of MDD was documented fully for only 1 of 18 patients (5.6 percent) who were said to have MDD. The same documentation omission error was made in every case. The diagnosis of MDD requires, first, that one or more episodes of major depressive episode (MDE) be documented and, second, that alternative explanations of the MDE (e.g., schizo-affective disorder, normal bereavement) be ruled out. In all but one case in the paper condition in which the clinician stated that the patient had MDD, alternative explanations for MDE were not ruled out. Although the paper checklist allowed the clinician to assert the diagnosis of MDD even though rule-out conditions for MDE had not been documented, the CaseWalker required the clinician to evaluate the presence or absence of those conditions when MDE was established. Thus, failure to document the rule-out criteria never occurred with the CaseWalker. The 100 percent rate of accurately documenting the diagnosis in the CaseWalker condition is a very significant achievement from a quality assurance perspective.

Limitations

The major limitation of the study is that the differential effect of the two reminder systems on clinical outcomes was not assessed. In this regard, this study is no different from 76 percent of randomized controlled trials of computer reminder systems,⁵ but it is a major limitation nonetheless.

A major clinical objective of the VHA clinical practice guideline for MDD was to increase the probability that true cases of MDD are diagnosed. Whether the two reminder systems were equally effective in accomplishing this objective depends on whether one assumes the diagnosis of MDD was made correctly in the paper checklist group even though it wasn't documented thoroughly. If that assumption is valid, then the reminder systems did not differ in clinical effectiveness, because the reported incidence of MDD was virtually the same in both arms of the study. However, the validity of the diagnoses given in each arm of the study was not independently established, so it is not known whether the paper checklist resulted in more false-positive diagnoses or just poor documentation of valid diagnoses.

The other limitation of the study is that it was conducted in only one setting with only four clinicians. Thus, the generalizability of the findings to other clinical settings may be limited.

Conclusions

This study compared the effectiveness of a computer reminder system and a paper checklist in supporting the implementation of two primary objectives of a clinical practice guideline for MDD; specifically, to increase screening for mood disorder and to improve the documentation of the diagnosis of MDD. For both objectives, the computer reminder system was shown to be more effective than the manual reminder system. Thus, the present findings replicate the findings of Tape and Campbell¹⁸ as well as those of Frame et al.¹⁹ under a very different set of clinical conditions.

This study was submitted to the Department of Medical Informatics at the University of Utah School of Medicine by the first author in partial fulfillment of the requirements of an MS degree. Development of the CaseWalker software was a team effort by Dale Cannon, PhD, Jeff Sells, PhD, and Robert Feldman.

References ■

1. Brown JB, Shye D, McFarland B. The paradox of guideline implementation: how AHCPR's depression guideline was adapted at Kaiser Permanente Northwest Region. *Jt Comm J Qual Improv.* 1995;21:5-21.
2. Walker RD, Howard MO, Lambert MD, Suchinsky R. Medical practice guidelines. *West J Med.* 1994;161:39-44.
3. Grilli R, Lomas J. Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Med Care.* 1994;32:202-13.
4. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med.* 1989;321:1306-11.
5. Balas EA, Austin SM, Mitchell JA, Ewigmann BG, Bopp KD, Brown GD. The clinical value of computerized information services: a review of 98 randomized clinical trials. *Arch Fam Med.* 1996;5:271-8.
6. Hunt DA, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA.* 1998;280:1339-46.
7. Austin SM, Balas EA, Mitchell JA, Ewigman BG. Effect of physician reminders on preventive care: meta-analysis of randomized clinical trials. *Proc 18th Annu Symp Comput Appl Med Care.* 1994:121-4.
8. Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *J Am Med Inform Assoc.* 1996;3:399-409.
9. Madlon-Kay DJ. Improving the periodic health examination: use of a screening flow chart for patients and physicians. *J Fam Pract.* 1987;25:470-3.
10. Madlon-Kay DJ. Use of a structured encounter form to improve well-child care documentation. *Arch Fam Med.* 1998;7:480-3.
11. Schreiner DT, Petrusa ER, Rettie CS, Kluge RM. Improving compliance with preventive medicine procedures in a house staff training program. *South Med J.* 1988;81:1553-7.
12. Yarnall KS, Michener JL, Broadhead WE, Tse CK. Increasing compliance with mammography recommendations: health assessment forms. *J Fam Pract.* 1993;36:59-64.
13. Cheney C, Ramsdell JW. Effect of medical records' checklists on implementation of periodic health measures. *Am J Med.* 1987;83:129-36.
14. Prislun MD, Vandenbark MS, Clarkson QD. The impact of a health screening flow sheet on the performance and documentation of health screening procedures. *Fam Med.* 1986;18:290-2.
15. Shank JC, Powell T, Llewelyn J. A five-year demonstration project associated with improvement in physician health maintenance behavior. *Fam Med.* 1989;21:273-8.
16. Kohatsu ND, Cramer E, Bohnstedt N. Use of a clinician reminder system for screening mammography in a public health clinic. *Am J Prev Med.* 1994;10:348-52.
17. Cohen DI, Littenberg B, Wetzel C, Neuhauser D. Improving physician compliance with preventive medicine guidelines. *Med Care.* 1982;20:1040-5.
18. Tape TG, Campbell JR. Computerized medical records and preventive health care: success depends on many factors. *Am J Med.* 1993;94:619-25.
19. Frame PS, Zimmer JG, Werth PL, Hall WJ, Eberly SW. Computer-based vs. manual health maintenance tracking: a controlled trial. *Arch Fam Med.* 1994;3:581-8.
20. Veterans Health Administration. Clinical Guideline for Major Depressive Disorder (MDD), MDD with Posttraumatic Stress Disorder (PTSD), and MDD with Substance Abuse (SA). Washington, DC: Department of Veterans Affairs, 1996.
21. American Psychiatric Association. Diagnostic and Statistical

- Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Press, 1994.
22. Fairbank JA, Schlenger WE, Saigh PA, Davidson JRT. An epidemiologic profile of post-traumatic stress disorder: prevalence, comorbidity, and risk factors. In: Friedman MJ, Charney DS, Deutch AY (eds). *Neurobiological and Clinical Consequences of Stress: From Normal Adaptation to PTSD*. Philadelphia, Pa: Lippincott-Raven, 1995:415–27.
 23. Wilkinson L. *Systat 8.0 Statistics*. Chicago, Ill: SPSS Inc, 1998.
 24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
 25. Bornstein M, Cohen J. *Statistical Power Analysis: A Computer Program*. Hillsdale, NJ: Lawrence Erlbaum, 1988.
 26. Chambers CV, Balaban DJ, Carlson BL, Grasberger DM. The effect of microcomputer-generated reminders on influenza vaccination rates in a university-based family practice center. *J Am Board Fam Pract*. 1991;4:19–26.