A Controlled Time-Series Trial of Clinical Reminders: Using Computerized Firm Systems to Make Quality Improvement Research a Routine Part of Mainstream Practice

Harold I. Goldberg, William E. Neighbor, Allen D. Cheadle, Scott D. Ramsey, Paula Diehr, and Ed Gore

Objective. To explore the feasibility of conducting unobtrusive interventional research in community practice settings by integrating firm-system techniques with time-series analysis of relational-repository data.

Study Setting. A satellite teaching clinic divided into two similar, but geographically separated, primary care group practices called firms. One firm was selected by chance to receive the study intervention. Forty-two providers and 2,655 patients participated.

Study Design. A nonrandomized controlled trial of computer-generated preventive reminders. Net effects were determined by quantitatively combining population-level data from parallel experimental and control interrupted time series extending over two-month baseline and intervention periods.

Data Collection. Mean rates at which mammography, colorectal cancer screening, and cholesterol testing were performed on patients due to receive each maneuver at clinic visits were the trial's outcome measures.

Principal Findings. Mammography performance increased on the experimental firm by 154 percent (0.24 versus 0.61, $p = .03$). No effect on fecal occult blood testing was observed. Cholesterol ordering decreased on both the experimental (0.18 versus 0.11, $p = .02$) and control firms (0.13 versus 0.07, $p = .03$) coincident with national guidelines retreating from recommending screening for young adults. A traditional uncontrolled interrupted time-series design would have incorrectly attributed the experimental-firm decrease to the introduction of reminders. The combined analysis properly indicated that no net prompting effect had occurred, as the difference between firms in cholesterol testing remained stochastically stable over time (0.05 versus 0.04, $p = .75$). A logistic-regression analysis applied to individual-level data produced equivalent findings. The trial incurred no supplementary data collection costs.

Conclusions. The apparent validity and practicability of our reminder implementation study should encourage others to develop computerized firm systems capable of conducting controlled time-series trials.

Key Words. Quality assurance, controlled clinical trials, firm system, time series, reminder system

---

If the quality and cost-effectiveness of the nation's healthcare system is to be improved, there is growing agreement that both clinical and administrative practice will need to become more evidence based. Less consensus exists regarding ways for all of the required evidence to be continuously updated as new drugs, procedures, and delivery options come on board. The costs of multicenter randomized controlled trials (RCTs), which remain the gold standard for evaluation of interventions, are becoming prohibitive. The fact that these studies are usually conducted on highly selected samples at tertiary referral centers engenders continuing debate regarding their generalizability to more typical patients in the community. The concern is that, apart from cost, sole reliance on RCTs would squander the opportunity to use the vast cumulative evidence being generated every day in mainstream clinical settings to identify best practices (Moses 1995).

Accordingly, this report is concerned with two methodologies that have been recently advanced as potential alternatives to the use of traditional RCTs. The first is an experimental-design variant known as "firm-system" research (Cebul 1991). Here, large inpatient services and/or outpatient clinics are divided into equivalent subunits, that is, firms, to which providers and patients are randomized in an ongoing process. By introducing changes on some firms, but not others, controlled trials of interventions such as group practice arrangements, intravenous therapy teams, and screening for alcoholism have been performed at low cost (Goldberg, Cohen, Hershey, et al. 1987; Tomford, Hershey, McLaren, et al. 1984; Goldberg, Mullen, Ries, et al. 1991). However,

---

This research was supported by grants from the Foundation for Health Care Quality (#5010-03920) and the National Library of Medicine (#5-608-LM05620-05).

Authors, all from the University of Washington, are Harold I. Goldberg, M.D., Associate Professor, Department of Medicine; William E. Neighbor, M.D., Associate Professor, Department of Family Medicine; Allen D. Cheadle, Ph.D., Research Associate Professor, Department of Health Services; Scott D. Ramsey, M.D., Ph.D., Assistant Professor, Department of Medicine; Paula Diehr, Ph.D., Professor, Department of Biostatistics; and Ed Gore, Ph.D., Senior Computer Specialist, Medical Centers Information Systems. Address all correspondence to Harold I. Goldberg, M.D., Harborview Medical Center, 325 Ninth Avenue, Box 359780, Seattle, WA 98104. This article, submitted to *Health Services Research* on July 9, 1998, was revised and accepted for publication on March 8, 1999.

although more than 25 percent of Veterans Affairs medical centers reported operating firm systems in 1992, their use has generally been otherwise limited to academic settings (Landefeld and Aucott 1995). It appears that within academia, the mandate to conduct rigorous research has overridden concerns expressed in the private sector that restricting choice of primary provider to the members of the firm to which the patient has been randomly assigned is simply not marketable.

Increasing attention has also been paid to the examination of aggregated population-level data over time known as "time-series analysis" (Orwin 1997). Popular in the social sciences, where controlled designs are not often feasible, these methods are now appearing in healthcare journals. For example, time series have been used to determine if ingestion-related fatalities decreased following the advent of child-resistant drug packaging—or if quarterly CABG mortality rates had decreased in response to a regional quality improvement program (Rodgers 1996; O'Connor, Plume, Olmstead, et al. 1996). However, even such "clinical" interrupted time-series experiments (ITSEs) often have not included comparison groups. When comparison groups are included, differences between ITSEs are examined qualitatively, but their data have generally not been quantitatively combined and subjected to tests of statistical inference. True control populations, where chance assignment to the experimental or control condition is made explicitly for prospective comparison, have not been employed. The absence of comparison or control groups leaves authors responsible for convincing readers that effects were not due to unrecognized, concurrent secular changes. Given the proliferation of fast-changing financial arrangements and clinical guidelines now affecting health outcomes, this can sometimes be a difficult task.

We speculated that these two methodologies might be synergistically integrated. In settings where ongoing randomization had been deemed impracticable, time-series analysis might provide historical control for the variability naturally exhibited between similar, but nonequivalent, firms. Conversely, co-situated firms might naturally provide for the kinds of quasi-experimental controlled comparisons that would strengthen time-series designs. To examine these issues, we took advantage of the introduction of a computerized reminder system into an academic family practice clinic to conduct an impromptu implementation trial. A recent meta-analysis of 16 RCTs concluded that the overall odds ratio for the effect of reminders on preventive-maneuver performance was 1.77 (Shea, DuMouchel, and Bahamonde 1996). Relative improvements ranged from none for cervical cancer screening, to over 300 percent for immunizations, varying with factors such as physician intention,

patient refusal, and the ease with which maneuvers could be obtained. We chose to study preventive reminders so that reader attention might remain focused on the research design issues involved rather than on the particular content of the intervention being tested. We reasoned that an intervention whose overall usefulness had already been amply demonstrated would be less distracting than one whose efficacy was still in doubt (Dietrich, Carney, Winchell, et al. 1997).

The clinic involved was chosen because we thought its operational arrangement would be widely generalizable. Like academic firm systems, and many large private sector clinics, the facility was divided into geographically separated subunits for the sake of efficiency. More typical of community practices, however, new patients presenting without provider preferences were assigned to primary physicians solely on the basis of appointment availability without the use of any kind of formal randomization scheme. Consequently, we undertook to (1) demonstrate the feasibility of quantitatively combining the information contained in parallel experimental and control ITSEs in order to measure the net effects of representative preventive reminders; (2) check the internal validity of this approach by comparing results with those obtained from applying logistic regression analysis to individual-level data from the same experience; and (3) discuss the advantages and disadvantages of such "controlled time-series trials" (CTSTs), especially regarding their potential conduct in mainstream practice settings that store claims and clinical data in computerized repositories.

METHODS

Study Setting

The Family Medical Center (FMC) at the University of Washington (UW) provides care to 7,700 patients in a satellite facility a mile west of the medical school campus. The clinic is geographically divided into two firms, each with its own personnel, examination rooms, and waiting area. Each firm is staffed by three small teams of two to three faculty members, a representative of each of the three residency class years, and a fulltime physician assistant. The three teams on either side of the clinic are supported by dedicated nurses, medical assistants, and receptionists. Because a provider from each team is present in the clinic at all times, cross coverage for unscheduled visits is generally provided by team members. Assignment of new residents to teams is made nonrandomly, without knowledge of personal characteristics other than the

conscious attempt to evenly distribute women residents. For this study, one firm was assigned by coin flip to receive the reminder intervention, and the other firm served as the control group. Characteristics of the 42 study providers are shown in Table 1.

The Clinical Data Repository

The facilities of the UW Academic Medical Centers (AMCs) include the only tertiary care hospital, level-one trauma center, and medical school available in the five contiguous "WWAMI" states of Washington, Wyoming, Alaska, Montana, and Idaho. Consistent with this regional mission, construction of a large relational repository, the Medical Information Networked Database (MIND), was undertaken in 1989. Its goal was to make both clinical and reference information available in real time to providers in the UW's far-flung referral base. Originally, MIND employed interfaces with legacy registration, billing, pharmacy, laboratory, radiology, pathology, and transcription computing systems to generate patient records that were viewable on the AMCs' local area network. In 1995, clinical informaticists began to collaboratively design a graphical HTML "front end" to the MIND repository. Called MINDscape, this user interface enabled repository content to be accessed over the Internet using standard browsers such as Netscape Navigator or Microsoft's Internet Explorer. The repository currently stores the medical records of 404,000 patients (Goldberg, Tarczy-Hornoch, Stephens, et al. 1998).

The Reminder System

The Clinical Reminder and Outcomes System (CROS) was the first decision support application written to run against the MIND database. Based on age, sex, and diagnoses, CROS prompts for the performance of indicated preventive and chronic disease processes and the collection of both physiological and functional outcome measures. The program acts as a population

Table 1: Provider Characteristics

Characteristic	*Experimental (n = 21)*	*Control (n = 21)*	p-*Value*
Age, mean years	38.2	35.5	.38
Sex, percent male	42.9	33.3	.52
Race, percent white	85.7	85.7	1.00
Teaching status, percent faculty	47.6	47.6	1.00
Time in practice, mean years	10.6	9.1	.62
Number of patients seen, mean/week	13.0	15.8	.22

monitor, preprocessing the current status of all primary care patients on all reminders each evening so that this information can be stored and displayed the following day. However, because no display terminals were located in examination rooms at the time of the implementation trial, CROS output was initially made available only in text version. A printed one-page sheet was placed on top of the clinic chart of each patient visiting the experimental firm. Subsequent to the trial's conclusion, output was also made available on-line. Both the printed and on-line versions of reminders for the current MINDscape "test patient" used for demonstrations are shown in Figure 1.

Study Design

The study was conducted between July 1 and November 30, 1996, employing a pretest, posttest quasi-experimental design. Results were compared between experimental and control firms over two-month baseline and intervention periods. These periods were separated by a one-month wash-in period during which operation of the reminder system was tested and experimental providers were briefed on its use at team meetings. Because CROS did not recognize new patients registering at their first encounter, the study included those 2,655 established patients, ages 18–75, who visited during the baseline or intervention period and who had also made at least one clinic visit during the previous two years. Patient characteristics are shown in Table 2. As examples, we chose to report the effect of reminders on three preventive maneuvers: yearly mammograms on women ages 50–60 and every two years between the ages of 60 and 75; screening for colorectal cancer with fecal occult blood cards on men and women ages 50–75 every two years, and; cholesterol determinations every five years on men and women ages 18–65. These were selected because they respectively represented the smallest ($n = 153$), the median ($n = 432$), and the largest subgroup of patients ($n = 948$) due to receive any individual maneuver. We wanted to ensure that our methodology was tested across the range of sample sizes observed. The study protocol was approved by the UW's Institutional Review Board.

Statistical Analysis

To compare physician and patient characteristics, we employed t-tests for continuous variables and contingency table analysis for categorical variables. Two distinct methods were applied to analyzing results, a traditional pre-post comparison using logistic regression and a controlled time-series analysis. The basic unit of analysis for both methods was the "opportunity," defined

Figure 1: Printed and On-line Versions of Reminders for the Current MINDscape Test Patient

```
FAMILY MEDICAL CENTER     CL 14 (880)                          Patient Profile Report For:
NEIGHBOR, W                                                          TESTMCIS, PTIGNORE
                                                                              U6999999

                                PATIENT DEMOGRAPHICS

Name: TESTMCIS, PTIGNORE
ID:   U6999999                                  Address: PO BOX 1002
Race:                          Sex: F                    BLDG 4
Born: 01-FEB-1979                                        SEATTLE, WA  98118
SSN:  333-33-3333                               Phone:   HOME: 555-1234
Language:                                                WORK: 555-8987
PCP:  NEIGHBOR, W (000794)                      Inpt Insurance:  CARENET
Clinic: FAMILY MEDICAL CENTER     CL 14 (880)   Outpt Insurance: SEBB UNIFORM PLAN


==[ PCP ASSIGNMENT REVIEW ]=====================================================

Please correct assigned PCP and clinic (service center), if needed, in the spaces below:

 Corrected PCP: _________________ (_______)   Corrected Clinic /: _ _____________ ____
                                                      Service Center:

==[ REMINDERS FOR PROVIDER ]====================================================

                                        Services DUE:
                                          Min                                Corrected  Not
Condition        Service                  Freq  Last Done  Due On    Last Done  UW    D/C
---------------  ------------------ -----  ----  ---------  --------  ---------- -  ---
Diabetes         Glycated hemoglobin      6 mo    09/06/96  03/06/97  _____________[ ]  [ ]
Diabetes         Retinal Exam             1 yr    09/06/96  09/06/97  _____________[ ]  [ ]
Diabetes         Urine protein            1 yr    09/06/96  09/06/97  _____________[ ]  [ ]
Diabetes         Foot exam                1 yr    09/06/96  09/06/97  _____________[ ]  [ ]


                                    Services NOT DUE (could be due soon):
                                          Min                                Corrected  Not
Condition          Service                Freq  Last Done  Due On    Last Done  UW    D/C
-----------------  ---------------- -------  ----  ---------  --------  ---------- ---  ---

Cervical cancer    Pap Smear              3 yr    06/25/96  06/25/99  _____________[ ]  [ ]
Diabetes           Cholesterol            3 yr    09/06/96  09/06/99  _____________[ ]  [ ]
```

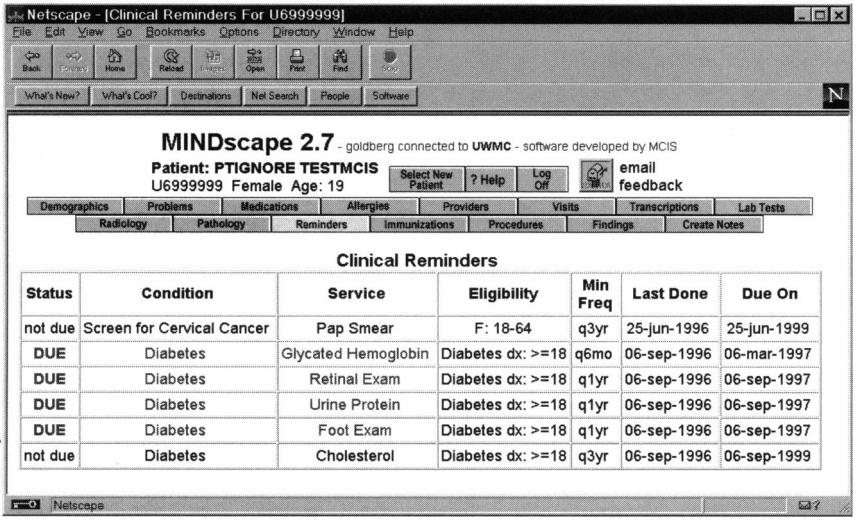


Table 2: Patient Characteristics

Characteristic	*Experimental (n = 1,433)*	*Control (n = 1,222)*	p-*Value*
Age, mean years	42.9	43.0	.86
Sex, percent male	33.7	35.4	.37
Race, percent white	79.1	78.6	.79
Marital status, percent married	49.1	51.5	.21
Payer, percent			
Commercial	71.2	70.9	.95
Medicaid	8.9	9.1	
Medicare	11.0	11.5	
Self-pay	2.4	2.7	
Other	6.5	5.9	
Reminder-related diagnoses, percent with			
Breast cancer	1.1	1.1	.95
Colon cancer	0.1	0.3	.13
Ischemic heart disease	5.7	6.5	.42

as a visit where the patient was due to have a preventive maneuver done (McConnochie and Roghmann 1992). A patient was considered due, if on the date of the visit there was no record in the repository to show that the indicated maneuver had been performed within the recommended interval. An opportunity was defined as being "converted" if the indicated maneuver was performed within ten days of the index visit for cholesterol and fecal occult blood tests, and, because advance scheduling was required, within 60 days for mammography.

Tests for statistical significance first used logistic regression analysis. Opportunity conversion rates for each reminder, that is, the percentage of opportunities converted, were examined over the entire baseline and intervention periods for both experimental and control firm patients. The dependent variable in the regressions was set equal to one if an opportunity was converted and zero if no maneuver was performed. The independent variables were dummies for time (baseline = 0, intervention = 1) and firm (control = 0, experimental = 1). The significance of the intervention effect was captured using a dummy variable for the interaction between time and firm. Because patients may have made multiple visits, and because patients were assigned along with their primary physicians to experimental or control status, the potential for clustering effects existed at both levels. These effects might have resulted in an underestimation of variance, which could have led to an overestimation of statistical significance. To obviate the need for separate

cluster sample analyses (Donner and Klar 1994), we controlled for patient-level clustering by selecting one visit per patient at random. Physician-level clustering was found not to be of significance when examined by repeating the analysis using a statistical procedure (SUDAAN) specifically designed to adjust for such clustering in the computation of variances.

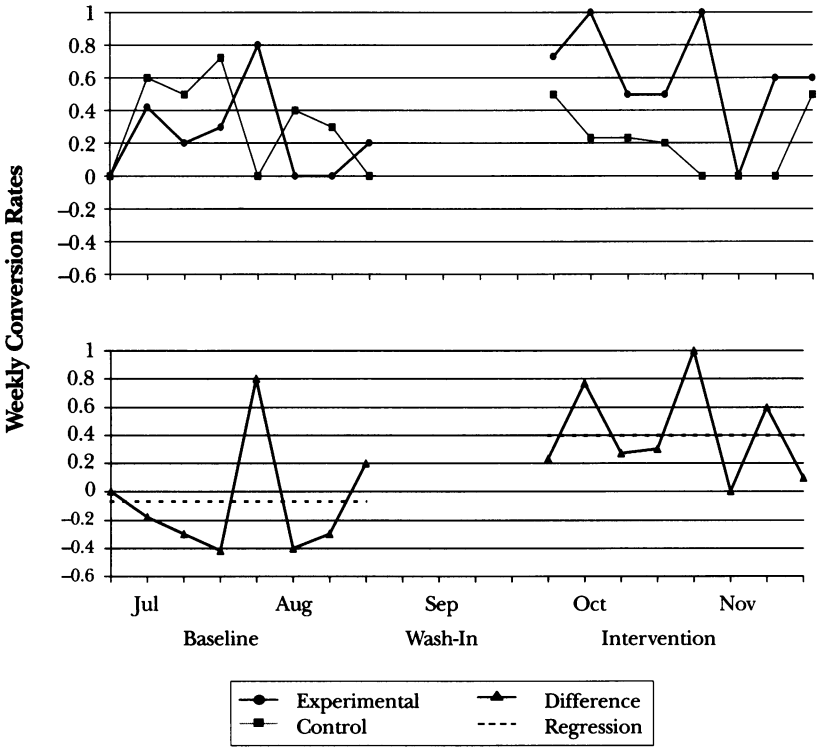
The time-series analysis used the same basic data as the logistic regression evaluation—including the same random visit for patients with multiple visits—but the data were aggregated and analyzed differently. Weekly conversion rates were generated and plotted as time series with eight observations during both the baseline and intervention periods for both firms. For example, if a firm experienced ten visits during a given week by patients due to have mammograms and the procedure was performed within 60 days on three of them, the opportunity conversion rate for that observation was 0.30 or 30 percent. To estimate the net effect of each of the reminder interventions, its control time series was subtracted from its experimental time series. A least-squares regression was then fitted to this "difference" time series that included a dummy variable for time (baseline = 0, intervention = 1) to test whether mean differences in conversion rates between experimental and control firms had significantly changed. The experiences of both firms were then similarly analyzed as separate ITSEs in order to determine their individual contributions to the net effect. Time-series regressions were estimated using Econometric Views software (Quantitative Micro Software, Irvine, CA). Correlograms and Durbin-Watson statistics were computed to test for serial correlation. In no case was the serial correlation either large or statistically significant, so no ARIMA adjustments (e.g., including lagged terms) were made when fitting the regressions.

Assuming reminders exerted an immediate effect, the regression model included only an intercept term. In the absence of serial correlation, the model yields results equivalent to those obtained from applying a *t*-test to the eight baseline and eight intervention observations in the difference time series. However, because regression fitting could also be applied to a comparison of slopes before and after implementing interventions expected to exert their effects gradually, it is presented as a more generalizable method for analyzing time-series data in firm trials. All tests of significance employed an alpha level of 0.05.

RESULTS

Time-series plots demonstrating the effect of reminders on mammography performance are included as Figure 2. Weekly conversion rates for the

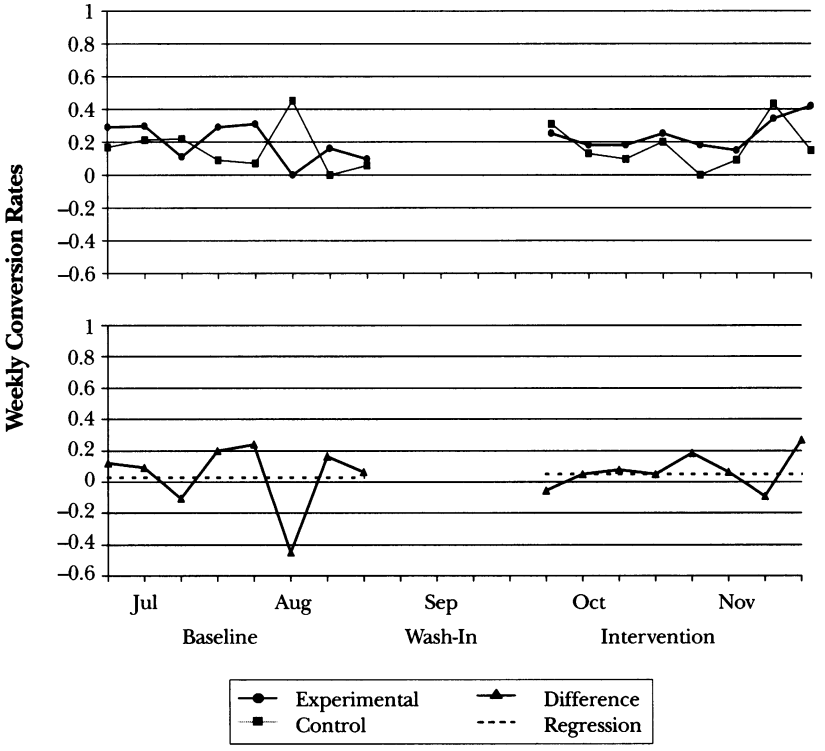
Figure 2: Weekly Conversion Rates for Mammography Are Plotted as Experimental and Control-Firm Interrupted Time Series; a Least-Squares Regression Is Fitted to the Time Series Representing Their Difference



experimental and control firms, as well as the difference between them, are shown. On average, rates on the experimental firm exhibited an absolute increase from 0.24 to 0.61 between the baseline and intervention periods, a relative increase of 154 percent ($p = .03$). In contrast, performance on the control firm was not affected (0.31 versus 0.21, $p = .44$). The regression line fitted to the difference time series indicated that the net effect of mammography reminders on mean conversion rates was an absolute increase of 0.47 (−0.07 versus 0.40, $p = .02$).

Results regarding colorectal cancer screening are shown in Figure 3. Fecal occult blood testing did not significantly change over the baseline and intervention periods on either the experimental firm (0.20 versus 0.25,

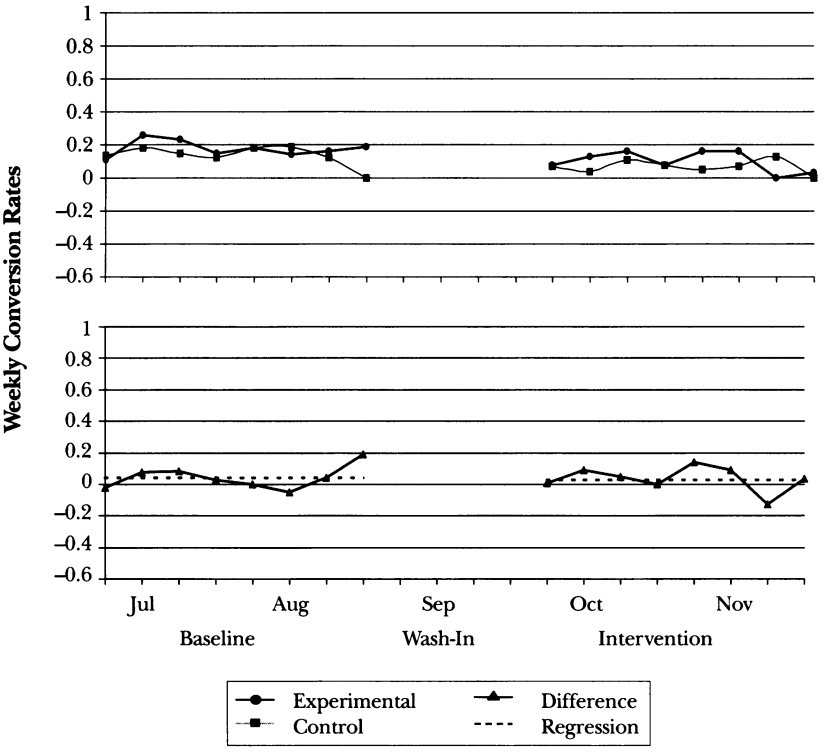
Figure 3: Weekly Conversion Rates for Fecal Occult Blood Are Plotted as Experimental and Control-Firm Interrupted Time Series; a Least-Squares Regression Is Fitted to the Time Series Representing Their Difference



$p = .33$) or the control firm (0.16 versus 0.18, $p = .73$). The difference time series confirmed the absence of a net reminder effect, as mean conversion rates were stable over time (0.04 versus 0.07, $p = .77$). This lack of effect may have represented physician uncertainty, given that current evidence is inadequate to determine whether fecal occult blood testing or sigmoidoscopy is the preferred screening modality (Frame, Berg, and Woolf 1997).

Cholesterol screening activity is depicted in Figure 4. Modest decreases in mean conversion rates were exhibited by both the experimental (0.18 versus 0.11, $p = .02$) and the control firms (0.13 versus 0.07, $p = .03$). The difference time series, however, indicated that reminders exerted no net effect on behavior, as mean rates were unchanged over the baseline and

Figure 4: Weekly Conversion Rates for Cholesterol Testing Are Plotted as Experimental and Control-Firm Interrupted Time Series; a Least-Squares Regression Is Fitted to the Time Series Representing Their Difference



intervention periods (0.05 versus 0.04, $p = .75$). This example illustrates the importance of including control groups when applying time-series techniques to the evaluation of clinical interventions. The experimental ITSE alone would have incorrectly attributed the observed decrease to the introduction of reminders. Instead, we believe the decrease on both firms was more likely the result of a secular shift in opinion. Both the U.S. Preventive Services Task Force and the National Committee for Quality Assurance had moved to exclude young adults at low risk for ischemic heart disease from their cholesterol screening guidelines (U.S. Preventive Services Task Force 1996). These changes became widely known and were discussed during the trial's tenure. A conscious determination was made to delay

reprogramming of the cholesterol reminder until after the trial had been completed.

The logistic regression analysis also determined that the introduction of reminders had significantly improved mammography, but not either the fecal occult blood or cholesterol conversion rates. This agreement indicated that the process of aggregating patient-level data had not resulted in meaningful information loss.

DISCUSSION

We sought to explore the feasibility of integrating firm-system techniques with time-series analysis of relational-repository data in a way that would facilitate the conduct of unobtrusive interventional research as a routine part of mainstream practice. Our CTST produced results that appeared valid on face, and agreed with those obtained via the application of logistic regression analysis to the same experience. We believe, however, that the advantages of the CTST design stem from its practicability. Given an increasingly competitive healthcare market, we cannot expect community physicians to perform uncompensated controlled trials that either hinder new patient recruitment, disrupt clinic flow and attendant productivity, or incur more than minimal data collection costs. Consequently, the reported trial did not affect patient enrollment or assignment. Staff in neither firm were required to differentially track or treat experimental versus control patients as is required by research designs involving ad hoc randomization. All data involved had been collected as part of routine operation, obviating any need for supplementary chart abstraction or survey completion.

Indeed, the intervention's incremental implementation—from the initial testing by several physicians during the wash-in period, to half of the clinic during the trial, to full deployment after the trial's conclusion—is one that might have occurred anyway as a change-management strategy had an institutional decision been made simply to install a clinical reminder system. Thus, it could be argued that the trial provided valuable evidence on the local effectiveness, or lack of same, of various reminders at a marginal cost of the data analysis alone. However, beyond confirming the local utility of approaches evaluated elsewhere, we would propose that CTSTs be used proactively. Suppose, for example, large health systems routinely conducted CTSTs to pretest promising health services interventions of uncertain utility at a small number of their clinic sites. Innovations determined to be cost-effective could then be confidently adopted. Assuming that some proportion

of these trials would prove negative, the savings from having prevented an expensive institution-wide deployment would be considerable.

CTSTs are also inherently graphical, a feature that lends itself well to longitudinal quality improvement efforts. Sizable changes in regression-line intercept or slope can be intuitively understood even by clinicians who are not statistically sophisticated. For example, the FMC has now embarked on a two-year CTST aimed at increasing the percentage of patients with diabetes who can maintain near-normal levels of glycemic control. One of the interventions will be to distribute time-series printouts to experimental firm physicians to provide ongoing feedback on how well they are doing.

The obvious disadvantage of the CTST design is that it represents a methodological "step downward" in the hierarchy of causal inference. With RCTs, randomization of adequate numbers generates prognostically equivalent experimental and control groups at baseline. Thus, any difference that subsequently develops is causally attributable to the intervention being evaluated. In contrast, CTSTs acknowledge that groups may differ at baseline but assume that this difference should be stochastically stable over time. Therefore, any significant deviation in the difference between groups that subsequently develops is causally attributable to the intervention of interest. Although it would be interesting to speculate on the degree to which baseline nonequivalences are "accounted for" in the differencing process, it is conceivable that any nonequivalence might interact with the intervention to produce differential responses in the experimental and control groups.

This is why we would still recommend that inter-firm differences be consciously minimized, even though the measured characteristics of our nonrandomly assigned providers and patients did not significantly differ between firms. As is done in traditional firm systems, assigning new providers and staff via randomization with replacement raises no marketing objection. Even when offered unfettered choice of primary provider, many new patients nonetheless continue to defer to the discretion of appointment staff. Such patients can be randomly assigned as well. As is done with all RCTs, baseline characteristics should be compared as part of the conduct of all CTSTs so that statistical adjustment can be applied if warranted.

Finally, there remains the issue of dissemination. As did the FMC, we believe that many mainstream clinics may already meet the dual prerequisites of CTST conduct, namely that facilities are divided into separate firms and that longitudinal data are readily accessible via computerization. The health organizational literature now suggests that small primary care groups are more likely to optimize care costs and patient satisfaction than are larger,

impersonal, and more complex practice arrangements (Barr 1995). At Group Health Cooperative of Puget Sound, the largest HMO in Washington State, primary care teams have been called "pods," a reference to the small groupings of marine mammals beloved in the Pacific Northwest. Growth in the use of clinical data repositories and fully interactive electronic medical records is also now being chronicled (Ornstein 1997).

In conclusion, given the vast amount of evidence that will be required to continuously improve the quality and cost-effectiveness of the nation's healthcare system, it is likely that both traditional and alternative research designs will need to be recruited to the effort. Accordingly, we believe that the apparent validity and practicability of our reminder implementation study should encourage others to explore the development of computerized firm systems capable of conducting CTSTs in their own clinical settings.

ACKNOWLEDGMENT

The authors would like to thank Mary Mullen, project coordinator, and Jim Hoath, Lee Fouche, and Natasha Manion for their help in respectively developing MIND, MINDscape, and CROS. We are also indebted to Cheryl Tabachnick for her invaluable secretarial assistance.

REFERENCES

Barr, D. A. 1995. "The Effects of Organizational Structure on Primary Care Outcomes Under Managed Care." *Annals of Internal Medicine* 122 (5): 353–59.

Cebul, R. D. 1991. "Randomized, Controlled Trials Using the Metro Firm System." *Medical Care* 29 (7, Supplement): JS9–JS18.

Dietrich, A. J., P. A. Carney, C. W. Winchell, C. H. Sox, and S. C. Reed. 1997. "An Office Systems Approach to Cancer Prevention in Primary Care." *Cancer Practice* 5 (6): 375–81.

Donner, A., and N. Klar. 1994. "Methods for Comparing Event Rates in Intervention Studies When the Unit of Allocation is a Cluster." *American Journal of Epidemiology* 140 (3): 279–89.

Frame, P. S., A. O. Berg, and S. Woolf. 1997. "U.S. Preventive Services Task Force: Highlights of the 1996 Report." *American Family Physician* 55 (2): 567–76.

Goldberg, H. I., D. I. Cohen, C. O. Hershey, I-L. Hsiue, D. K. Porter, and C. E. McLaren. 1987. "A Randomized Controlled Trial of Academic Group Practice: Improving the Operation of the Medicine Clinic." *Journal of the American Medical Association* 257 (15): 1051–55.

Goldberg, H. I., M. Mullen, R. K. Ries, B. M. Psaty, and B. P. Ruch. 1991. "Alcohol Counseling in a General Medicine Clinic: A Randomized Controlled Trial of

Strategies to Improve Referral and Show Rates." *Medical Care* 29 (7, Supplement): JS49–JS56.

Goldberg, H. I., P. Tarczy-Hornoch, K. Stephens, E. B. Larson, and J. P. LoGerfo. 1998. "Internet Access to Patients' Records." *Lancet* 351 (9118): 1811.

Landefeld, C. S., and J. Aucott. 1995. "Improving Primary Care in Academic Medical Centers: The Role of Firm Systems." *Medical Care* 33 (3): 311–14.

McConnochie, K. M., and K. J. Roghmann. 1992. "Immunization Opportunities Missed Among Urban Poor Children." *Pediatrics* 89 (6, Part 1): 1019–26.

Moses, L. E. 1995. "Measuring Effects Without Randomized Trials? Options, Problems, Challenges." *Medical Care* 33 (4, Supplement): AS8–AS14.

O'Connor, G. T., S. K. Plume, E. M. Olmstead, J. R. Morton, C. T. Maloney, W. C. Nugent, F. Hernandez, Jr., R. Clough, B. J. Leavitt, L. H. Coffin, C. A. Marrin, D. Wennberg, J. D. Berkmeyer, D. C. Charlesworth, D. J. Malenka, H. B. Quinton, and J. F. Kasper. 1996. "A Regional Intervention to Improve the Hospital Mortality Associated with Coronary Artery Bypass Graft Surgery." *Journal of the American Medical Association* 275 (11): 841–46.

Ornstein, S. M. 1997. "Electronic Medical Records in Family Practice: The Time Is Now." *Journal of Family Practice* 44 (1): 45–48.

Orwin, R. G. 1997. "Twenty-One Years Old and Counting: The Interrupted Time Series Comes of Age." In *Evaluation for the 21st Century: A Handbook,* edited by E. Chelimsky and W. R. Shadish, pp. 443–65. London: SAGE Publications.

Rodgers, G. B. 1996. "The Safety Effects of Child-Resistant Packaging for Oral Prescription Drugs: Two Decades of Experience." *Journal of the American Medical Association* 275 (21): 1661–65.

Shea, S., W. DuMouchel, and L. Bahamonde. 1996. "A Meta-Analysis of 16 Randomized Controlled Trials to Evaluate Computer-Based Clinical Reminder Systems for Preventive Care in the Ambulatory Setting." *Journal of the American Medical Information Association* 3 (6): 399–409.

Tomford, J. W., C. O. Hershey, C. E. McLaren, D. K. Porter, and D. I. Cohen. 1984. "Intravenous Therapy Team and Peripheral Venous Catheter–Associated Complications: A Prospective Controlled Study." *Archives of Internal Medicine* 144 (6): 1191–94.

U.S. Preventive Services Task Force. 1996. *Guide to Clinical Preventive Services: Report of the U.S. Preventive Services Task Force, 2d ed.* Baltimore, MD: Williams & Wilkins.