Effect of a Controlled Feedback Intervention on Laboratory Test Ordering by Community Physicians

Peter S. Bunting1,2* and Carl van Walraven3,4,5

Background: Most studies of interventions to reduce laboratory test utilization have occurred in academic hospital settings, used historical controls, or have had short postintervention follow-up. Interventions with the greatest impact use multiple approaches, are repeated regularly, include comparisons with physician peers, and have a personal approach. We determined whether laboratory test utilization by community physicians could be reduced by a multifaceted program of education and feedback.

Methods: We identified 200 physicians who ordered the largest number of common laboratory tests during 1 year in a nonhospital, commercial community (reference) laboratory. They were assigned to intervention and control groups (100 each). Intervention physicians were visited individually up to three times by laboratory representatives over a 2-year period. At each visit, educational material and the physician's personal laboratory test utilization data were presented and discussed briefly in general terms, with the latter compared with utilization data for the physician's peers. Overall test utilization rates 1 year before, during, and 2 years after the intervention were measured using population-based databases. Time-series analysis was used to determine the effect of the intervention on laboratory test utilization.

Results: The two groups began with similar test utilization: control group, 4.06 × 10^6 tests in 1.48 × 10^6 visits (2.73 tests/visit); intervention group, 3.90 × 10^6 tests in 1.41 × 10^6 visits (2.77 tests/visit). During the 2-year intervention, intention-to-treat analysis showed that utilization decreased significantly in the intervention group compared with the controls [relative reduction of 7.9% (P < 0.0001); absolute reduction of 0.22 tests/visit (95% confidence interval, 0.20–0.24)]. This difference persisted until the end of study observation, or more than 2 years after the intervention ended.

Conclusion: A multifaceted education and feedback strategy can significantly and persistently decrease laboratory utilization by practicing community physicians.

There have been numerous attempts over the past four decades to modify the laboratory test ordering behavior of physicians, mostly because of perceived or real overutilization (1–8). Many of these studies did not use a control group, and most control groups that were used were historical (8) and were subject to biases such as the Hawthorne effect (improved behavior while under observation) (9). The majority of these studies were carried out in hospital settings. A structured literature review in the Cochrane database (10) concluded that education and feedback achieved variable success.

We have identified nine studies (11–19) that examined the effects of various interventions on laboratory utilization by physicians in the out-of-hospital outpatient setting. These community studies were very different from our own. Most of them were not aimed at general test utilization, but rather at specific components such as cancer screening, thyroid function test guidelines, or Papanicoleau (PAP) smears. Two studies within the Harvard Community Health Plan were associated with a highly academic facility. The approach of the Winkens group is resource-intensive (17–19), and beyond the scope of most nonacademic communities.

Overviews of test utilization have concluded that mod-

1 Gamma-Dynacare Medical Laboratories, Brampton, Ontario, Canada.
2 Department of Pathology and Laboratory Medicine, The Ottawa Hospital and University of Ottawa, Ottawa, Ontario, Canada.
3 Departments of Medicine and Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.
4 Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa, Ontario, Canada.
5 Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada.

*Address correspondence to this author at: Division of Biochemistry, Department of Pathology and Laboratory Medicine, The Ottawa Hospital, Civic Campus, 1053 Carling Ave., Ottawa, Ontario, K1Y 4E9 Canada. E-mail psbunting@ottawahospital.on.ca.

Received July 29, 2003; accepted November 14, 2003.
Previously published online at DOI: 10.1373/clinchem.2003.025098

© 2004 American Association for Clinical Chemistry
Most outcomes are achieved with utilization audits; some studies using presentation of test costs to the ordering physician have had an impact (20). Success is most likely if more than one strategy is used at a time (8, 21), if feedback is close to the time of decision-making, and if clinicians are willing to review their practice (22). Success is more likely if the intervention is ongoing (23). Peer comparison can have a significant effect on changing practice patterns (24). Unfortunately, most studies have found that test-ordering patterns return to previous rates as soon as the intervention is stopped (25, 26). It is therefore important to monitor postintervention behavior for prolonged periods to determine the true effect of an intervention on utilization. It is also important to have a nonhistorical control group.

We conducted a clinical trial to determine whether a multifaceted education and feedback program could reduce the test-ordering behavior of physicians practicing in a community setting.

**Materials and Methods**

**Study Setting**

In Ontario, laboratory tests for physicians in out-of-hospital private practice are done in commercial community laboratories, which are similar to privately owned reference laboratories in the US. Physicians must request laboratory tests on an Ontario Health Insurance Plan (OHIP) form, developed jointly by laboratory professionals and the Ministry of Health and Long Term Care. Unlike the US, there are no test panels; each test must be requested individually by a combination of checking off a limited list of tests and writing other tests by hand. The community laboratory is reimbursed for the majority of these routine laboratory tests by OHIP. For several years this reimbursement has been capitated. This means that OHIP payment to these community (reference) laboratories is fee-for-service up to a predetermined amount; there is no reimbursement beyond the capitated amount, irrespective of the volume of testing. This provides an incentive to laboratories to manage utilization. A similar capitation system applies to physician reimbursement.

Each test is identified by a unique laboratory test code (so-called L-code). The test-ordering transactions are stored in a central database, which was accessed for this study. Tests without an L-code are not reimbursed by OHIP, but are paid for by the patient. Sometimes the patient can be reimbursed for these by private health insurance plans. Tests without L-codes form a minority of all clinical laboratory tests.

**Physician Selection**

We selected the 200 physicians who ordered the greatest number of laboratory tests in 1997 from a single commercial community (nonhospital) laboratory. Utilization was based on total dollar charges for tests ordered and was not corrected for patient volume (dollars correlated very closely with number of tests ordered). These physicians were ranked from highest to lowest in terms of utilization and alternately placed in an intervention group and a control group (100 in each). This approach was used to stratify for dollar volume because there were large differences between some physicians. No attempt was made to equalize the number of specialists and generalists in each group. Physician gender, practice specialty, and date of graduation were determined from the laboratory's data-base and from the Canadian Medical Directory (27). Charges were obtained from the Ontario Schedule of Benefits for Physician Services (28).

**Intervention**

This was a multifaceted education and feedback intervention (Table 1), consisting of an introductory meeting

---

**Table 1. Dates and types of activities monitored during the study period.**

<table>
<thead>
<tr>
<th>Period</th>
<th>Dates</th>
<th>Description of activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January 1, 1997–October 30, 1998</td>
<td>Preintervention monitoring of both groups</td>
</tr>
<tr>
<td>2</td>
<td>November 1, 1998–February 28, 1999</td>
<td>Start of intervention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Introduction to the program and individual interviews with intervention physicians</td>
</tr>
<tr>
<td>3</td>
<td>March 1, 1999–May 31, 1999</td>
<td>First feedback of physician utilization data to intervention group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>General notice to all physicians (study group and others) about the program</td>
</tr>
<tr>
<td>4</td>
<td>June 1, 1999–September 30, 1999</td>
<td>Post-feedback monitoring</td>
</tr>
<tr>
<td>5a</td>
<td>October 1, 1999–February 28, 2000</td>
<td>Second feedback visit with presentation of physician utilization data to intervention group</td>
</tr>
<tr>
<td>5b</td>
<td>June 1, 2000–September 30, 2000</td>
<td>Third feedback visit to intervention group; included publication on introduction interview&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>6</td>
<td>March 1, 2001–April 30, 2001</td>
<td>End of intervention period</td>
</tr>
</tbody>
</table>

<sup>a</sup> See Bunting (29).
followed by three subsequent feedback visits, and took place over 2 years. The first meeting occurred between November 1998 and February 1999 and represented the start of the intervention. Each intervention physician was hand-delivered an introductory letter, which stated that he or she was one of the top laboratory users in the region. The letter invited them to participate in a program to decrease their test volume. At the same time, a Client Service Representative (CSR) introduced the physician to the program and requested his or her participation. Physicians were already familiar with the CSRs, based on previous working relationships. Each physician was shown his or her own personal monthly laboratory test utilization rates from one community (reference) laboratory. These data were compared with the means of the intervention and control groups as well as with a group of 100 middle-use physicians. Data were presented as number of tests per physician per month, number of tests per requisition, and dollar charges of each of these.

The second meeting occurred between March and May 1999; at this meeting physicians were presented their utilization data as in the first meeting, but this time plotted graphically. Graphs were similar in style to Fig. 1. If the CSR could not meet with a physician after multiple attempts, the data were left in the physician’s mailbox. The third meeting occurred between October 1999 and February 2000, at which time each physician was again given graphic representations of his or her laboratory utilization data as in the second meeting. In June 2000, intervention physicians were mailed a study summarizing the baseline interviews (29). In the final component of the intervention (in March 2001), a letter was sent to the intervention physicians outlining the group’s utilization data to that time compared with controls and indicating that their utilization would continue to be monitored. No personalized graphs were given at that time. This represented the end of the intervention period. The first meeting lasted ~20–30 min; the second and third lasted 5–10 min, depending on physician availability.

During the study, control physicians were aware of a utilization management program in general terms, but they were not provided any information on their own test use.

**DATA COLLECTION, ANALYSIS, AND STATISTICS**

Physician identifiers were encrypted to link with the Ontario Physician Services Database. This database records claims for all insured (i.e., funded by the provincial government) assessments of Ontario physicians as well as all insured laboratory tests conducted by non-hospital-based community laboratories such as the one in which our study was conducted. Each Physician Services Database claim records the date of the service and the physician who conducted the patient assessment or ordered the test. From January 1, 1997, to April 30, 2002, we retrieved reimbursement claims for all patient visits to study physicians in out-of-hospital clinics (fee code starting with “A”) or nursing homes (fee code starting with “W”), as well as claims for laboratory tests (fee code starting with “L”). For each week of the observation period, we calculated laboratory test utilization rates for the intervention and control groups as total number of laboratory tests divided by the total number of patient assessments.

The difference in weekly rates was used to create a time series. Interventional autoregressive, integrated, moving average (ARIMA) time-series modeling was used to determine whether rate differences between groups changed after the intervention started (30, 31). Time-series analysis was used because data with a temporal sequence are often autocorrelated (i.e., the value at time $t$ is affected by the value at time $t - 1$). Therefore, the error terms of these observations are not independent, making simple regression and standard inferential testing inappropriate. Time-series analysis models autocorrelation in temporally sequenced data, permitting traditional regression and inferential testing. In our model, the feedback intervention was represented by a dummy series (0 representing preintervention and 1 representing postintervention) that was cross-correlated with the rate difference series. The autocorrelation function plots and partial autocorrelation function plots were reviewed to determine the presence of moving average or autoregressive processes, respectively. If the $t$-ratio for the intervention parameter in the best fitting model had a two-sided $P$ value $< 0.05$, the intervention was considered to be associated with a significant change in laboratory utilization. The final ARIMA model fit the data well, with Q6 and Q12 statistics of 0.65 and 0.75, respectively.

Analysis was intention-to-treat, i.e., data for all physicians initially assigned to the intervention and control groups were analyzed as such (although some physicians did not receive all interventions).

Data collected during January through October 1998 (preintervention, just before the introductory visit) are identified as period 1, data collected between the introductory visit and the first feedback visit relate to period 2; periods 3 through 5 include the data for subsequent feedback periods. Period 6 includes postintervention data collected until April 2003 (Table 1).

**Results**

The characteristics of the physicians in both groups at the start of the intervention are listed in Table 2. Distribution of practice specialty, year of graduation, and gender were all statistically equivalent ($P > 0.05$).

There was some physician attrition in both groups. After the end of the second feedback visit, five physicians had left the intervention group and had received only one to two of the intervention visits. After the final contact in March 2001, seven other physicians had left the intervention group and had received only two to three of the intervention visits.

The number of tests per visit for the control and
intervention groups during the study, as well as their difference, are shown in Fig. 1. The two groups had very similar test utilization rates during the preintervention stage (period 1). Test utilization decreased sharply in the intervention group, especially during the early stages of the intervention. The control group also experienced a somewhat smaller decrease in utilization after the intervention started, which could be a spillover effect of the intervention. During the postintervention stage (period 6), both control and intervention groups showed increased numbers of tests per visit toward the baseline, but the difference between groups remained. The time-series model indicated an absolute mean decrease of 0.22 tests/visit (95% confidence interval, 0.197–0.241) after the intervention started. The baseline mean before the intervention was 2.8 tests/visit for the intervention group, indicating that the intervention was associated with a 7.9% (P < 0.0001) relative reduction in laboratory utilization. We also noted a nonsignificant decrease compared with baseline in test utilization for the control group during the intervention, amounting to 0.14 tests/visit (95% confidence interval, −0.02 to 0.30).

Table 2. Physician characteristics in control and intervention groups.

<table>
<thead>
<tr>
<th>Practice specialty</th>
<th>Year of graduation</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP ab CCFP Other c</td>
<td>&lt;1960</td>
<td>1960s</td>
</tr>
<tr>
<td>Control physicians</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>Intervention physicians</td>
<td>67</td>
<td>17</td>
</tr>
</tbody>
</table>

* x² = 1.379 (P = 0.502) for practice specialty; 0.660 (P = 0.956) for year of graduation; and 0.268 (P = 0.875) for gender (all nonsignificant).

b GP, general practitioner; CCFP, Canadian College of Family Physicians or equivalent specialty in family medicine.

c Physicians with other specialties (control; intervention): obstetrics and gynecology (n = 1; n = 4), rheumatology (n = 2; n = 1), nephrology (n = 2; n = 2), urology (n = 2; n = 0), internal medicine (n = 1; n = 2), pediatrics (n = 0; n = 1), general surgery (n = 0; n = 1), and medical biochemistry (n = 1; n = 0).

The mean utilization rates for both groups during each period of the study are shown in Table 3.

**Discussion**

We found that a multifaceted education and feedback strategy significantly and persistently decreased laboratory utilization among practicing community physicians. There were some general trends in utilization in both the intervention and control groups: downward during the active part of the intervention (weeks 76–150 in Fig. 1), followed by a slight trend upward thereafter. These variations in both groups may be the result of a spillover effect resulting from intervention and control physicians talking to one another. The control group was aware of a utilization project in general terms but received no utilization data of any sort. However, these trends could also be the result of other unknown effects. There is a slight drift downward in the rate difference between control and intervention groups toward the end of the observation period (week 100 onward). These observations highlight the need for a concurrent control group, which in this case experienced a nonsignificant decrease from baseline.

Fig. 1. Effect of intervention on laboratory utilization by community physicians.

This plot presents laboratory utilization rates (expressed as the number of tests per visit) for the control (thin solid line) and intervention groups (thick solid line) between January 1, 1997, and March 31, 2002. The dotted line represents the difference between control and intervention groups. Intro, introduction.
Without this control group, the effect of the intervention could have been overestimated. These observations also highlight the need for randomization in physician assignment to minimize bias.

There were no obvious differences in physician characteristics in the control and intervention groups that might explain the observed difference in utilization (Table 2). The absolute (0.22 tests/visit) and relative (7.9%) sizes of the effects of the intervention may seem small, but when multiplied through a large database of physicians they could amount to a large savings. The annual capitated payment for all community laboratory testing in Ontario in 2003–2004 is $526 million (32). If the top 20% of users of laboratory tests were to be targeted by a similar intervention province-wide, there could be a reduction in laboratory test costs of approximately $8 million (Canadian).

The effect of this intervention was not entirely expected because the literature reports variable success with such interventions (1–10). Indeed, a recent randomized controlled multicenter trial, which targeted specific patient disorders and assessed whether appropriate testing guidelines were being followed, showed only a modest effect of intervention (33). The success of intervention in the present study could stem from several factors. Information was transmitted to the physicians in various forms, as suggested by Solomon et al. (8), Valenstein (7), and Grimshaw et al. (21). Our education and feedback information was personalized as extensively as possible and included a form of peer pressure, which could have a strong influence over physician behavior (14, 17–19). Finally, there were multiple visits over the approximately 2 years of the intervention. A regular reiteration is thought by many to be important to the persistent success of any intervention that attempts to influence physician behavior (3).

The most surprising aspect of the study, however, is the continued observance of an effect even after the intervention was halted. This runs counter to the literature, which suggests that when most interventions have been stopped, the effect disappears. This difference may be the result of continued postintervention visits by CSRs to these same physicians about other unrelated laboratory matters, which could serve as a reminder of laboratory test utilization. It is also possible that with further passage of time the observed difference could lessen or disappear.

This study took place in the setting of a capitated reimbursement environment in Ontario. Thus, both laboratories and physicians were aware of cost-containment issues. Nonetheless, we believe that this study is relevant to other provinces in Canada and to health maintenance organizations in the US and elsewhere. By identifying high-use physicians in a practice area and with use of a relatively small amount of resources (computer time, graphic reports, CSR time for brief discussions of feedback materials with physicians), a similar reduction in tests ordered should be feasible.

Our study has some limitations. The number of physicians in each group was relatively small and may not be generalizable to other physicians because of sampling issues. Indeed, the selection of the largest users of laboratory tests from a single laboratory was done to ensure the largest effect. It is highly likely that more moderate users of the laboratory would show a smaller effect. Whether the relative benefit of the intervention would remain the same is unknown. The literature suggests that continued feedback, if reinstituted, could continue to maintain the difference. We were able to monitor L-coded tests only (see the section on study setting in the Materials and Methods). Examples of some non-L-coded tests are specialized antibody and hormone tests, most metals, and some drugs. However, L-coded tests form the vast majority of the tests done, and there is no reason to suspect differences between intervention and control physicians in this regard. The physician selection was pseudo-randomized in alternate fashion from a ranked list. There is no reason to believe that this would lead to any permanent bias in outcome. We did not keep an accurate account of time spent by each CSR with each physician to identify the actual cost of the intervention. Dollar values for tests were charges rather than actual costs. Finally, this study did not monitor test appropriateness, but only the volume of testing corrected for number of patient visits. There are situations in which more testing improves

### Table 3. Total test utilization and patient visits for study physicians.*

<table>
<thead>
<tr>
<th>Study period</th>
<th>Control physicians, per time period</th>
<th>Intervention physicians, per time period</th>
<th>Control minus intervention (tests/visit), %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tests, × 10^6</td>
<td>Visits, × 10^6</td>
<td>Tests/Visit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.055</td>
<td>1.484</td>
<td>2.73</td>
</tr>
<tr>
<td>2</td>
<td>0.666</td>
<td>0.285</td>
<td>2.58</td>
</tr>
<tr>
<td>3</td>
<td>0.507</td>
<td>0.201</td>
<td>2.52</td>
</tr>
<tr>
<td>4</td>
<td>0.614</td>
<td>0.281</td>
<td>2.45</td>
</tr>
<tr>
<td>5</td>
<td>1.441</td>
<td>0.607</td>
<td>2.37</td>
</tr>
<tr>
<td>6</td>
<td>3.614</td>
<td>1.396</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tests, × 10^6</td>
<td>Visits, × 10^6</td>
<td>Tests/Visit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.897</td>
<td>1.408</td>
<td>2.77</td>
</tr>
<tr>
<td>2</td>
<td>0.593</td>
<td>0.254</td>
<td>2.33</td>
</tr>
<tr>
<td>3</td>
<td>0.450</td>
<td>0.190</td>
<td>2.36</td>
</tr>
<tr>
<td>4</td>
<td>0.548</td>
<td>0.245</td>
<td>2.24</td>
</tr>
<tr>
<td>5</td>
<td>1.271</td>
<td>0.512</td>
<td>2.18</td>
</tr>
<tr>
<td>6</td>
<td>3.227</td>
<td>1.334</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control minus intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Direct difference^b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normalized^c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Numbers for tests and visits differ greatly because periods refer to different lengths of time in the study.

\(^b\) Difference between control and intervention tests per visit, divided by control, expressed as a percentage.

\(^c\) Same as in footnote \(^a\), but normalized for the groups to be equalized at the start of study, to better assess the size of the intervention effect.
patient health, for example, when more screening is needed to detect undiagnosed diabetes mellitus (34) and in the case of more rigorous application of cancer-screening guidelines (12, 16).

In conclusion, this study showed a statistically significant effect of a simple feedback and education intervention on the ordering of laboratory tests by high-volume community physicians. It is the first such study that we are aware of in Canada and was carried out under “real life” circumstances and in the midst of other activities relating to laboratory testing. Analysis was intention-to-treat, thus undervaluing the intervention because some physicians did not receive all visits. The effect was maintained for the 2 years monitored after the intervention period. The observation of variations in utilization in both the intervention and control groups during the study highlighted the need for a concurrent control group. The observation of a significant effect has implications for potential savings by community laboratories (if their test volume is above their capitated payment) and/or the various agencies that fund them.

References