Objectives: To describe the variation among physicians in test ordering when caring for children with gastroenteritis and to explore the effect of hospital charge information on such variation.

Design: Prospective, nonmasked, observational study and controlled trial of price information.

Setting: Urban, university-affiliated pediatric emergency department.

Participants: Pediatric emergency medicine faculty (n=10) and fellows (n=5).

Methods: Test-ordering practices were reviewed during 3 periods: control, intervention, and washout. During the intervention period, test charge information was placed on patients' emergency department records. Telephone contact with families was initiated 7 days after care.

Results: We included 3198 visits. Individual physician mean test charges varied more than 2-fold during the control period (mean, $127; range, $82-$185). Based on their test charges (control period), physicians were assigned to the “high” (n=8) or “low” (n=7) test user group. Differences in mean charges in high vs low test users during the control period ($144 vs $112) persisted in the intervention period ($80 vs $52; Mann-Whitney P = .01), as did rates of intravenous fluid use (20% vs 14% in both periods). Among the lowest-acuity patients, low test users exhibited greater price sensitivity (vs high users). Patients treated by low test users did not differ in improved condition (82% vs 86%) or family satisfaction (93% vs 92%); they had more unscheduled follow-up (25% vs 17%; P<.01), but were no more often admitted (5% vs 3%; P=.11).

Conclusions: Physicians varied in resource use when treating children with gastroenteritis. High and low test users were sensitive to price information. This intervention did not seem to compromise patient outcome.


Practice variation among physicians is well documented.1,2 This variation in part reflects uncertainty in medicine: although the number of clinical guidelines is increasing, in many specific clinical situations physicians determine the tests or interventions needed for their patients.3 Thus, clinicians develop individual styles of patient management based on their experiences, their training, and their temperament. Some physicians are comfortable ordering fewer tests, whereas others prefer the additional information made available by using more studies.

Inefficient use of diagnostic tests at academic institutions has been observed.4 This seems in part to be a result of the academic environment, where the variety of payer relationships and hospitalwide billing procedures effectively insulate the individual physician from the relative costs of diagnostic testing.5 In contrast, physicians in community practices face logistical and direct financial disincentives to ordering tests.6,7 Test-ordering behavior may also reflect resident training, most of which occurs in the academic setting, and the career track of many academic pediatric emergency medicine physicians, who remain in an academic setting after completing their training.

Although it is difficult to know the specific tests needed to provide high-quality medical care, it is clear that inefficient use of diagnostic studies contributes to health care costs.8,9 Providing test price information has been demonstrated to result in fewer tests being ordered.10,11 In a previously published work, Hampers et al13 described the general effect of price information in the emergency department (ED), without detailing inter-
physician variations. In this study, we describe the variation in test-ordering behavior in a single setting (the pediatric ED) among physicians with similar training (pediatric emergency medicine faculty and fellows) in managing a single disease (gastroenteritis). We also looked at the effects of the provision of test charge information on individual physician test-ordering practices and patient outcomes.

**METHODS**

This study was conducted in the ED of an urban, tertiary care pediatric hospital with an annual volume of 39000 patients. Board-certified pediatric emergency medicine physicians or fellows, working independently, supervised the care of all patients studied. Although pediatric, emergency medicine, and family practice house staff participated in the care of many patients, they had limited autonomous decision-making authority.

On arrival at the ED, patients were triaged by acuity of illness to 1 of 4 categories: emergent, urgent high, urgent low, and nonurgent. Between September 1, 1997, and March 31, 1998, a data form was attached to each patient chart at triage. The form asked physicians to identify patients aged 2 months to 10 years who had a complaint of vomiting, diarrhea, decreased oral intake, or fever. We defined gastroenteritis as vomiting or diarrhea, and we included all children with those complaints. Children with chronic illness (a history of immunosuppression or immunodeficiency, an inborn error of metabolism, or a ventriculoperitoneal shunt) were excluded.

Data were collected during 3 periods. During the control period (September 1, 1997, to December 31, 1997), physicians were asked to circle the tests they ordered on the data form. During the intervention period (January 1, 1998, to March 31, 1998), physicians were asked to circle the tests they ordered on a similar data form that included the hospital charge for each test. During the washout period (April 1-30, 1998), the forms used in the control period were again used. The control period was 1 month longer than the intervention period because of seasonal variation in the incidence of gastroenteritis; this allowed enrollment of a similar number of cases in each of these 2 periods. Hospital charges for individual tests did not change during the study.

One of the investigators (L.C.H.) or a research assistant reviewed the ED records after the visit to collect the following information: age, race or ethnicity (as recorded by registration personnel), insurance status, initial vital signs, diagnostic testing, use of intravenous fluids, disposition (admitted to the hospital or discharged from the ED), and attending or fellow supervising care. We calculated years in practice for supervising physicians as years since completion of residency.

During the control and intervention periods, patient families were contacted by telephone approximately 1 week after the ED visit. Using a structured interview, caretakers were asked to describe the child’s overall condition (better, same, or worse), whether they had taken the child to see a health care provider since the ED visit (scheduled or unscheduled and office or urgent care/ED), and whether the child had been admitted to the hospital. If the caretaker had been with the child at the time of the ED visit, they were asked to describe their overall satisfaction with their child’s care (very unsatisfied, somewhat unsatisfied, somewhat satisfied, very satisfied). Caretakers who spoke only Spanish were interviewed in Spanish. Because the washout period was designed only to estimate the effect of the intervention on test charges, telephone interviews were not conducted during that time.

Data were analyzed using statistical software (SPSS for Windows; SPSS Inc, Chicago, Ill.). For comparisons between groups of categorical data we used the χ² test; we used a 2-tailed t test for continuous variables. Test charges, which were not normally distributed, were compared between physicians and between periods using the Mann-Whitney test; 95% confidence intervals (CIs) for means are reported. Statistical significance was set at P<.05. The study was approved by the institutional review board of Children’s Memorial Hospital (Chicago).

**RESULTS**

Data were available for 1415 visits during the control period and 1429 visits during the intervention period. Review of daily ED records suggested that the data we report represented an estimated 90% of eligible cases; reasons for inappropriate exclusion included clerical error (no form was attached to the ED record) and physician error (the physician did not complete the form). The demographic and clinical characteristics of eligible patients who were not included in this study did not differ from those of patients who were included.

Each physician (n = 15) treated a mean (SD) of 94 (29) patients (range, 40-136) during the control period and 95 (31) patients (range, 44-159) during the intervention period.

The 15 participating physicians exhibited wide variation in their clinical approaches. Physician-specific mean test charges during the control period ranged from $82.40 to $185.29 (overall mean, $126.66; median, $74.00). Based on this variation in test charges, physicians were divided into 2 groups: “high” test users (6 faculty physicians and 2 fellows) and “low” test users (4 faculty physicians and 3 fellows). Mean (SD) duration of practice was 7 (5) years for high test users and 4 (1) years for low test users (Mann-Whitney P = .20). Mean test charges associated with high vs low test users during the control period were $143.84 (95% CI, $130.39-$157.28) vs $112.43 (95% CI, $101.32-$123.54) (Figure).

The demographic and clinical characteristics of children cared for by high and low test users are given in Table 1. There were no significant differences between groups in race or ethnicity, insurance status, triage category, or clinical characteristics. We also compared the case-mix of patients within each period; there were no significant differences between high and low test users in patient distribution by triage category during either
period ($\chi^2$, control period: $P = .40$; and intervention period, $P = .98$).

**Table 2** gives test charges for patients cared for by high and low users during the control and intervention periods stratified by triage category. Because emergent patients accounted for less than 0.5% of each group, they were combined with urgent high patients for ease of data analysis and presentation. Within triage categories, the charge differences between high and low test users were statistically significant for each category and within each period except for urgent high patients during the intervention period. Overall, the mean charges for tests were 48% (95% CI, 39%-56%) lower during the period when physicians were provided with price information. The biggest declines were observed among low test users (−53% overall) and for patients who were triaged as nonurgent (−58% overall and −67% for low test users).

Among patients admitted to the hospital (17% during the control period and 11% during the intervention period), there were no significant differences in test charges between high and low test users during either period (control and intervention, Mann-Whitney, $P = .56$ for both), and there were no significant decreases in charges between periods (Mann-Whitney, $P = .09$). Among patients not admitted to the hospital, there were significant mean charge differences between periods (control period, $85.58$; intervention period, $36.88$; Mann-Whitney $P < .01$) and within periods between high and low test users.

More patients had no tests ordered during the intervention period (66% vs 41%; $\chi^2$ $P < .01$) (**Table 3**). The differences between high and low test users were significant for each test and for no tests during the control period. These differences were less marked during the intervention period.

High test users administered intravenous fluids to 20% of patients during the control and intervention periods; low test users gave intravenous fluids to 14% of their patients during each period. Patients treated vs not treated with intravenous fluids had higher test charges during the control period (chemistry charges, $72.80$ vs $5.34$; and total charges, $91.57$ vs $303.76$). Among patients treated with intravenous fluids by high test users, charges declined significantly between the control and intervention periods (chemistry charges, $77.00$ vs $47.80$; and total charges, $314.56$ vs $223.96$; Mann-Whitney $P < .01$ for each). Among patients treated with intravenous fluids by low test users, charges also declined significantly between periods (chemistry charges, $67.91$ vs $43.86$; and total charges, $291.17$ vs $176.50$; Mann-Whitney $P < .01$ for each).

We successfully obtained telephone follow-up information from 556 (39%) of 1415 families during the control period and 599 (42%) of 1429 families during the intervention period; 555 (48%) of these 1151 had been cared for by high test users and 600 (52%) had been treated by low test users. The percentage of children who were “better” was similar in practice style groups (86% for high users vs 82% for low users; $\chi^2$ $P = .20$), as was the proportion of caretakers who reported that they were “somewhat satisfied” or “very satisfied” with their ED visit (92% vs 93%; $P = .40$). Among children treated by low test users, there was more unscheduled follow-up (25% vs 17%; $P < .01$), mainly with primary care providers. Unscheduled returns to the ED occurred in 12% of patients cared for by low test users and 9% cared for by high test users ($P = .08$). The proportion of children who were admitted to the hospital during follow-up was similar (3% of patients treated by high test users and 5% of those treated by low test users; $P = .11$).

A total of 354 patients were included in the brief washout period. Patients of high users (n = 135) had charges similar to those observed during the intervention period (mean, $80.79$; Mann-Whitney $P = .99$); low users (n = 219) had a modest increase in charges (+11%), which was not statistically significant (mean, $58.71$; Mann-Whitney $P = .35$).

<table>
<thead>
<tr>
<th>Table 1. Demographic and Clinical Characteristics of Children Cared for by 8 High and 7 Low Test Users*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td>Race or ethnicity, No. (%)</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Not available</td>
</tr>
<tr>
<td>Insurance, No. (%)</td>
</tr>
<tr>
<td>Commercial</td>
</tr>
<tr>
<td>Medicaid</td>
</tr>
<tr>
<td>Uninsured</td>
</tr>
<tr>
<td>Triage category, No. (%)</td>
</tr>
<tr>
<td>Emergent</td>
</tr>
<tr>
<td>Urgent high</td>
</tr>
<tr>
<td>Urgent low</td>
</tr>
<tr>
<td>Nonurgent</td>
</tr>
<tr>
<td>Clinical characteristics, mean (SD)</td>
</tr>
<tr>
<td>Age, mo</td>
</tr>
<tr>
<td>Temperature, °C</td>
</tr>
<tr>
<td>Heart rate, beats/min</td>
</tr>
<tr>
<td>Respiratory rate, breaths/min</td>
</tr>
</tbody>
</table>

*Because of rounding, percentages may not all sum to 100.

**Comment**

There was enormous variation among physicians in tests ordered for pediatric patients with gastroenteritis, suggesting that some practiced in a resource-intensive manner and that others provided more efficient care. Variation in patient acuity does not explain these practice differences, and neither does family insurance status. As a result, average patient charges within triage categories vary 2-fold or greater. Providing price information was associated with a significant decrease in the number of tests ordered, reduced overall charges, and reduced some of the variation among physicians in test-ordering charges, particularly among visits triaged as urgent high.

Hampers et al previously described the general effect of price information in the ED, without detailing interphysician variations. The focus of the present study was to explore the implications of this intervention on the practice styles of faculty and fellows. In addition to
reporting striking individual practice variation, we believe that this analysis provides information about the receptiveness of members of a group of academic practitioners to educational interventions designed to alter their behaviors.

Results of the present analysis suggest that high and low test users are sensitive to price information. Among higher-acuity patients, high test users had greater declines than low test users, suggesting greater discretionary test use among high test users during the control period. During the intervention period, the mean charge difference between the high and low test users was less, suggesting a more consistent approach to test use in higher-acuity patients.

Overall test charges declined the greatest among less acute patients (–58% for nonurgent patients), suggesting that price information had the greatest impact on discretionary test ordering when physicians were treating children who were the least ill. Despite their lower utilization rates in the control period, low test users seemed to be even more sensitive to price information than were high users, further widening the gap between these 2 groups in the intervention period. It may be that the test-ordering practices of the low test users in part reflected a predisposition to price sensitivity. The high test users may have personality traits that made them less price sensitive, resulting in higher resource use in the control period and less sensitivity to charge information. This finding suggests an interesting paradox: price information may have the greatest impact on providers who are already practicing the most efficiently. Some of the variation in test ordering may reflect training environment: high test users, on average, had been in practice longer than low test users. Low test users may have been trained during a more “cost-conscious” time.

We do not believe that the resource-efficient physicians we observed did not obtain needed tests. More specifically, the data for the low test users suggested that rates of family satisfaction, reports that the child was better, and rates of hospital admission were similar to those observed among their higher-spending peers.
Although mean charges declined for patients within each stratum, there were no significant differences between the control and intervention periods in charges for patients who were admitted to the hospital. Although the methods used to estimate dehydration in children are imprecise and were not standardized for this study, we observed consistency among physicians across periods in the percentage of children treated with intravenous fluids. Thus, test use among children admitted to the hospital and treatment with intravenous fluids were not sensitive to test price information. Specific charge information was provided only for tests; it is possible that additional charge information for other interventions (intravenous fluid administration and hospital admission) would alter the results we present.

The washout period was brief and involved a limited number of patients. However, among high test users, overall charges during the washout period were similar to those in the intervention period, and low test users showed a modest increase. This suggests that short-term effects of the intervention on high and low test users persisted.

Other pediatric studies have observed differences in test ordering specific to practice settings and to training. There is clear variation among care providers in the tests obtained on febrile infants. Differences in the care of children with febrile seizures treated in community EDs vs university-affiliated children's hospitals have also been reported. Practice variation between physicians trained in pediatric emergency medicine and those trained in emergency medicine in the management of croup has been demonstrated. The present study, performed at a single institution among physicians who had completed or were in training to complete pediatric emergency medicine fellowships, suggests that individual behaviors also account for the significant variation in test-ordering practices.

Although gastroenteritis is not a high-cost disease, it is a high-volume disease, and the variation in charges becomes significant over time. There were clear differences in the use of intravenous fluids among high and low test users that was unexplained by case-mix differences. During the intervention period, fewer patients had “routine” tests ordered (serum electrolytes, complete blood cell count, urinalysis, and urine culture). A higher fraction of patients had no tests ordered. The challenge is that the “best” test-ordering strategy, which allows for only needed tests and maximizes efficiency for children with gastroenteritis, is not known.

High and low test users were sensitive to price information. Although these data suggest that information that caused efficient physicians to order even fewer tests resulted in satisfactory outcomes, additional study of this issue is warranted. Because most cases of gastroenteritis in infants and toddlers have a good prognosis, it is not surprising that the variation in testing we report did not seem to impact outcome or satisfaction.

It is difficult to completely separate the effect of attending physician test-ordering behavior from that of the house staff they supervised. Some of the house staff (postgraduate years 1 and 2) participated during both periods; others spent a single month during either the control or the intervention period in the ED. Because each physician (attending or fellow) supervised all house staff, it is unlikely that systematic differences between house staff explain the variation among supervising physicians. Although the fellows supervised house staff independent of the attending, it is possible that the attending had an impact on the tests that the fellow ordered. However, because each attending was present with each fellow at various times, it is unlikely that this explains the variation we report within each stratum and between periods.

It is also difficult to completely separate the effect of the incidence of gastroenteritis, which has seasonal variation, on attending physician test-ordering behavior. It is possible that physicians ordered fewer tests during the intervention period because gastroenteritis was more common then. Although this may account for some of the differences between periods, it is unlikely to explain the variation in charges between high and low test users during each period. Although our case-mix seems consistent in control and intervention periods, there may have been unrecognized differences in acuity, as the proportion of children admitted to the hospital was lower in the intervention period. However, as we reported in the results, significant differences in test charges persisted in a separate analysis of patients who were not admitted to the hospital.

Because there are no generally accepted outcome measures for children with gastroenteritis treated and then discharged from the ED, we chose to obtain information about family perception that the child was better, family satisfaction with the ED encounter, rates of hospital admission during follow-up, and rates of unscheduled follow-up with primary care providers or in the ED. Although we observed a higher rate of unscheduled follow-up when fewer tests were ordered, there were no differences in hospital admission rates or in parent satisfaction. The reasons for the higher rates of unscheduled follow-up are not clear but may include unmet parental expectations for tests or physician failure to obtain needed studies. We do not know whether diagnostic studies were obtained during follow-up visits. Follow-up visits contribute to overall health care costs; thus, further study is needed to address this observation. Nevertheless, we believe it is unlikely that failure to obtain studies contributed to significant morbidity, as parent satisfaction and hospital admission rates were similar for both groups.

Our study is limited by failure to obtain follow-up information on many patients. We used a fairly narrow time window, and many families provided incomplete contact information. Although this limits the generalizability of our results, there seem to be no systematic biases in follow-up rates between high and low test users, suggesting that the data we present are valid.

Gastroenteritis represents a narrow range of complaints, and we limited this study to healthy children. In addition, we defined gastroenteritis broadly as vomiting, diarrhea, or both, and it is possible that children with other diagnoses were included in the data we present. For most children, this illness is self-limited, and thus we were unlikely to observe poor outcomes. For this reason, much test ordering in the management of gastroenteritis is discretionary. It is clear that within a single
Data about physician-specific variation in test-ordering behavior and the effects of charge information are limited. We describe the variation in test use in a single ED among physicians with similar training in managing children with gastroenteritis. The data suggest that there was significant variation and that high and low test users were sensitive to price information. Low test use did not seem to be associated with poor outcomes.

Accepted for publication March 13, 2003.

This study was funded in part by a special project grant from the Ambulatory Pediatric Association, McLean, Va. We thank Robert R. Tanz, MD, for his thoughtful review of this manuscript.

Corresponding author: Elizabeth C. Powell, MD, MPH, Division of Pediatric Emergency Medicine, Children’s Memorial Hospital, Box 62, Chicago, IL 60614 (e-mail: epowell@northwestern.edu).

REFERENCES