LCC Session 18

CanMEDS Competency: Manager

Dr. W Liao, Dr. A. Burke, Dr. C Hillis

What will happen in this session?

Applicability to CanMEDS Roles:
Definition: As Managers, Pediatricians are integral participants in health care organizations, organizing sustainable practices, making decisions about allocating resources, and contributing to the effectiveness of the health care system.

Key and Enabling Competencies: Pediatricians are able to...
1. Participate in activities that contribute to the effectiveness of their health care organizations and systems
   1.2.2. The assessment of cost/benefit ratios of diagnostic and therapeutic interventions, cost-containment and efficacy, effectiveness and efficiency as they relate to quality assurance
   1.4. Describe principles of health care financing, including physician remuneration, budgeting and organizational funding

3. Allocate finite health care resources appropriately
   3.1. Recognize the importance of just and ethical allocation of health care resources, balancing effectiveness, efficiency and access with optimal patient care
   3.2. Apply evidence and management processes for cost appropriate care

Background:
Unnecessary diagnostic tests are a major source of healthcare waste
Health care providers have access to a wide array of diagnostic investigations to assist in patient care. Each investigation has variable utility depending upon the clinical context. Studies have found 27% of investigations ordered on the day of admission are avoidable, increasing to 63% on subsequent days¹. The Institute of Medicine estimates that one out of every three dollars spent on healthcare is wasted and that diagnostic testing is particularly inefficient².

Diagnostic tests are complex processes requiring considerable resources. Porters, nurses, ward clerks and lab technicians may be involved when a single investigation is ordered. Unnecessary testing generates unnecessary costs while delaying useful tests. Wasteful testing is also a concern with respect to patient safety. Iatrogenic injury may occur when false positive results lead to invasive investigations
or treatments. Reducing wasteful investigations has increased patient satisfaction in some clinical areas.

The majority of tests ordered on HHS Clinical Teaching Units do not advance patient care
The HHS Multi-year Operating Plan outlines the need to improve efficiency while maintaining quality. At present, investigation-ordering practices on HHS Clinical Teaching Units (CTUs) are not always informed, structured or efficient. Following an in-depth study of resident physician test ordering practices, independent clinician reviewers determined that 61% of diagnostic tests ordered locally were unnecessary. Commonly ordered investigations including complete blood count (CBC), urea and aspartate aminotransferase/alanine aminotransferase (AST/ALT), were unnecessary in more than 80% of cases.

The majority of tests ordered on HHS Clinical Teaching Units do not advance patient care
Assessment of the nature of errors made by resident physicians on HHS CTUs was assessed. In 53% of cases, repeat tests were ordered although the patient had not clinically changed and the prior result was normal. It was common for bloodwork to be ordered daily without any clinical question to answer. Investigations of this nature are unlikely to change diagnosis or management. Understanding the errors made in local practice provides a foundation to build the knowledge translation solution.

In addition to practice patterns, surveys of resident physician attitudes at HHS show cost-efficiency to be a marginal concern. 72% of residents never consider cost implications as part of their daily practice. Furthermore, 63% are completely unfamiliar with the costs of investigations they order frequently. These findings suggest that culture and attitude change needs to occur for gains in efficient diagnostic testing to be made. It has been shown physician value systems and clinical practice patterns can be impacted by effective educational programs.


Session # 1 – Large Group Exercise

In this session we will discuss some fundamentals of healthcare financing in Ontario. We will then present a framework by which test ordering currently occurs. As a large group we can brainstorm ways in which pediatrics residents can become more effective and efficient resource managers. Prior to the session please consider the tests you order on a daily basis, the motivations you have for ordering various diagnostics tests, and the barriers to better resource allocation.

Suggested Time 60 minutes.
Readings:

Physician Practice Variation in the Pediatric Emergency Department and Its Impact on Resource Use and Quality of Care

Shabnam Jain, MD,* Lisa K. Elon, MS, MPH,‡ Brent A. Johnson, PhD,†
Gary Frank, MD,§ and Michael DeGuzman, MPH§

Objective: To evaluate variation in case-mix adjusted resource use among pediatric emergency department (ED) physicians and its correlation with ED length of stay (LOS) and return rates.

Methods: Resource use patterns at 2 EDs for 36 academic physicians (163,669 patients at ED1 and 45 private physicians (289,199 patients at ED2) from 2003 to 2006 were abstracted for common laboratory tests, imaging studies, intravenous therapy (fluids/antibiotics), LOS and 72-hour return rate for discharged patients, and hospital admissions for all patients. Case-mix adjustment was based on triage acuity, diagnostic category, demographics, and temporal measures. Outcome measures: (1) adjusted overall resource use for ED1 and ED2 physicians and (2) observed-to-expected ratios for ED1 physicians.

Results: Case-mix adjusted hospital admission rates among physicians varied nearly 3-fold (6.3%–18%) for ED1 and 8-fold (2.5%–19.4%) for ED2. Intravenous therapy use varied 2-fold (4.9%–10.4%) at ED1 and 3-fold (3.6%–11.4%) at ED2. Emergency department 2 physicians had an almost 2-fold (10.9%–20.6%) variation in imaging use. Variation in head computed tomography use was 2-fold (1.1%–2.5%) at ED1 and 5-fold (0.9%–4.8%) at ED2. Physicians had longer than expected LOS if they had higher than expected use of laboratory tests (r = 0.41; 95% confidence interval [CI], 0.09–0.65; P < 0.05) and imaging (r = 0.48; 95% CI, 0.17–0.69; P < 0.01). Return rate was not significantly correlated with resource use in any category. Physicians with higher than expected use of laboratory tests had higher than expected use of imaging (r = 0.62; 95% CI, 0.36–0.78; P < 0.001), head computed tomography (r = 0.49; 95% CI, 0.19–0.70; P < 0.01), and intravenous therapy (r = 0.51; 95% CI, 0.20–0.71; P < 0.01).

Conclusions: Significant variation exists in physician use of common ED resources. Higher resource use was associated with increased LOS but did not reduce return to ED. Practice variation such as this may represent an opportunity to improve health care quality and decrease costs.

Key Words: practice variation, resource use, quality improvement

(Pediatr Emer Care 2010;26:902–908)

The impact of excessive variation in medical practice on the cost and the quality of care is currently under scrutiny, both in lay press and in the health care industry.2–7 Work by Wennberg et al8–13 and James14 has directed the health care system’s attention to variation in health care delivery in adults. In pediatrics, comparisons of tonsillectomy rates and asthma care between different geographic regions have shown significant practice variation.15,16 Other studies on practice variation in pediatrics have also been limited to the management of specific disease conditions, such as asthma, bronchiolitis, gastroenteritis, sinusitis, and febrile neonates.17–27 Some of these studies have compared practice patterns based on differences in type or level of training or differences in practice setting.17,18,28–38 The Agency for Healthcare Research and Quality recently reported that variation in health care delivery remains high and suggested that pediatric quality efforts should focus on transforming care over large areas of health, rather than specific disease entities.39,40 We are not aware of any published studies that have evaluated differences in practice between similarly trained physicians in the same practice setting who are influenced by similar practice variables such as patient population characteristics, capacity issues, resource availability, and temporal trends.

Several attributes make the emergency department (ED) an ideal setting to compare individual practitioners’ practice patterns. Patients present across a wide spectrum of conditions, making it possible to study practice patterns across a range of conditions. Rapid patient turnover and management by a single physician for a given ED visit make physician attribution more accurate in the ED. Furthermore, larger numbers of patients allow for adequate denominator size, and standardized triage systems allow for uniform adjustment for severity of illness. Finally, ED care is often criticized for excessive use of discretionary diagnostic and therapeutic interventions. Issues such as hospital admission rates and use of computed tomography (CT) lend themselves well to studying variation between individual practitioners in an ED.

The primary objectives of this study were to (1) evaluate the case-mix adjusted variation in physician resource use in 2 pediatric EDs and (2) examine the correlation between resource use, ED length of stay (LOS), and 72-hour ED return rate (RR). The secondary objective was to examine the correlation between physician resource use in one category (eg, laboratory tests or imaging) and use in other categories.

METHODS

The study was conducted in the emergency departments of 2 free-standing tertiary-care children’s hospitals, with more than 100,000 combined annual ED visits. The 2 EDs are in physically separate urban locations of a large city. Both EDs are staffed 24 hours by attending level physicians trained in pediatric emergency medicine or general pediatrics; most patients, except the rare patient in extremis, are seen by either type of physician. Patients are assigned in order of arrival to the first available attending physician, thus eliminating bias in patient selection. Emergency department 1 is staffed by academic faculty, whereas...
ED 2 is staffed by nonacademically affiliated, contracted attending physicians. All patients are supervised by an attending physician, who manages patients primarily or supervises trainees or midlevel providers; only attending physicians were included in the study. Both EDs have an electronic medical record (EMR) that tracks all orders with real-time electronic signature of the ordering physician.

Because the study involved only previously collected data that were completely deidentified and no interventions were made, the institutional review board exempted this study from institutional review board review.

From January 1, 2003 through December 31, 2006, 477,087 patient visits were recorded in an existing hospital-based data warehouse that houses a record of every ED visit to either hospital. The data warehouse holds the ED EMR and a mix of clinical and billing information systems.

**Patient Selection and Exclusions**

Patient visits with disposition of “left without being seen” (n = 4566), “left against medical advice” (n = 318), “EMR downtime” (n = 12), “discharged to law enforcement” (n = 54), “expired” (n = 91), and other miscellaneous dispositions (n = 578) were excluded. To remove bias from low numbers of patients, only physicians who saw at least 1000 patients during the study period were included; 15,561 patient visits attended by non-ED physician staff (who saw <1000 patients each in the study period) were excluded. Patients 19 years or older were excluded from all analyses (n = 3039). Our study population therefore has 452,868 patient visits: 163,669 visits at ED1 attended by 36 academic physicians and 289,199 visits at ED2 attended by 45 private practice physicians. Overall resource use patterns were evaluated for both ED1 and ED2. Practice patterns for physicians at ED1 were studied in detail at the individual provider level.

**Outcome Variables**

Admission was defined as disposition to inpatient or observation status, intensive care unit, operating room, or transfer to another facility. In the subset of patients who were discharged home, we examined use of 3 resource categories: (1) imaging—any use of noncontrast head CT, chest radiography, and abdominal radiography (head CT is reported as part of imaging and also independently); (2) laboratory tests—any use of basic metabolic panel, comprehensive metabolic panel, complete blood count, and rapid strep test; (3) intravenous (IV) therapy—any use of IV fluids or 2 commonly used IV antibiotics, ceftriaxone or cefotaxime. We also examined 2 recognized ED quality metrics: LOS, a measure of ED efficiency, defined as time from first examination by physician or midlevel provider to disposition home by a physician, and a patient’s return to the ED (RR) within 72 hours of a previous visit, a measure of adequacy of management at first visit.

Covariates included (a) demographic characteristics: sex, age (0–12 months, 13–59 months, 5–12 years, and 13–18 years), and ethnic group (1% refused to designate; missing for 7501 patients); (b) time of patient arrival by shift (7 A.M.–3 P.M., 3 P.M.–11 P.M., and 11 P.M.–7 A.M.), month, and year; (c) major diagnostic category based on International Classification of Diseases, Ninth Revision (ICD-9) codes; (d) severity of illness: triage acuity based on Emergency Severity Index (ESI) system of triage into 3 acuity categories (emergency [ESI levels 1 and 2], urgent [ESI level 3], and nonurgent [ESI levels 4 and 5]); Emergency Severity Index is a widely used ED triage algorithm that provides clinically relevant stratification of patients into 5 levels from 1 (most urgent) to 5 (least urgent) on the basis of acuity and resource needs.41

**Statistical Analysis**

Basic descriptive statistics were calculated for each ED, with median and range/interquartiles for continuous variables and percentages for categoric measures. We used standard statistical methods for comparing observed provider-level resource use rates to expected provider-level resource use rates based on an internal standard. Observed-to-expected (OE) ratios and their 95% confidence intervals (CIs) were estimated for each

| TABLE 1. Demographic and Clinical Characteristics of Pediatric ED Patients by Site (2003–2006) |
|-----------------|-----------------|-----------------|-----------------|
|                 | ED1             | ED2             |                 |
| Total visits    | 163,669         | 289,199         |                 |
| No. physicians  | 36              | 45              |                 |
| ED visits/patient, median (interquartile range) | 2 (1–5) | 2 (1–4) | |
| Male, %         | 54              | 54              |                 |
| Age group, %    |                 |                 |                 |
| 0–12 mo         | 21              | 22              |                 |
| 13–59 mo        | 37              | 38              |                 |
| 5–12 yr         | 30              | 29              |                 |
| 13–18 yr        | 12              | 11              |                 |
| Race/ethnicity, % |                |                 |                 |
| Asian           | 1               | 2               |                 |
| Black/African American | 68      | 24              |                 |
| Nonwhite Hispanic | 3             | 13              |                 |
| White, Hispanic | 2               | 10              |                 |
| White, non-Hispanic | 23           | 47              |                 |
| Other           | 3               | 4               |                 |
| Shift, %        |                 |                 |                 |
| 7 A.M.–3 P.M.   | 31              | 30              |                 |
| 3 P.M.–11 P.M.  | 49              | 49              |                 |
| 11 P.M.–7 A.M.  | 19              | 21              |                 |
| Triage level, % |                 |                 |                 |
| Nonurgent       | 53              | 53              |                 |
| Urgent          | 32              | 31              |                 |
| Emergent        | 15              | 16              |                 |
| Diagnostic categories: based on | ICD-9 codes, % |                 |                 |
| Ill-defined*    | 21              | 20              |                 |
| Respiratory     | 20              | 17              |                 |
| Injury          | 18              | 22              |                 |
| Nervous system/sense organs | 10     | 9               |                 |
| Digestive disease | 8          | 9               |                 |
| Infectious disease/parasites | 5    | 8               |                 |
| Skin disease    | 4               | 3               |                 |
| Endocrine/nutrition/metabolic/immunologic | 2 | 2 | |
| Blood disease   | 2               | 2               |                 |
| Genitourinary   | 2               | 2               |                 |
| Musculoskeletal system | 2  | 2               |                 |
| Poison/abuse    | 2               | 3               |                 |
| Other           | 4               | 4               |                 |

The sum of percentage values may not equal 100% because of rounding.

*Includes symptom-based diagnoses, for example, fever, seizure, cough, difficulty breathing, vomiting, and diarrhea (these 6 diagnoses accounted for nearly 75% of ill-defined category).
physician \((i = 1, 2, \ldots, 36)\) in ED1, using the following expression from DeLong et al.\(^4\)

\[
CI_i = \frac{1.96 \sqrt{\sum p_i(1-p_i)}}{\sum p_i}
\]

where \(p_i\) represents the predicted probability of a dichotomous outcome for the \(i\)th physician and \(j\)th patient visit (the dummy variable \(j\) ranges from 1 to the total number of patients seen by the \(i\)th physician in the study period). For dichotomous outcomes, predicted probabilities were calculated based on logistic regression models with the following independent variables: patient sex, age group, ethnic group, shift, triage acuity, month, year, ICD-9 category, and 3 interaction terms (triage by age group, triage by shift, and triage by ICD-9 category). These independent variables constitute the internal ED1 standard. Then, for each physician, the predicted probabilities were summed over patient \((\sum p_{ij})\) and interpreted as the expected use rate for the \(i\)th physician based on the internal ED1 standard. The observed dichotomous outcomes were summed similarly, and the ratio of sums was called the OE ratio.

To predict LOS, the concept of OE ratios was as previously discussed, with some modifications. Instead of logistic regression, for LOS, we used a generalized estimating equations modeling procedure with a gamma distribution, which accounts for the skewness in the empirical distribution of LOS. The predicted minutes of LOS were summed by physician, as were the observed minutes, and the ratio of sums formed the OE ratio. The 95% CI equation was

\[
CI_i = \frac{1.96 \sqrt{\sum (m_{ij} - M_{ij})^2}}{\sum m_{ij}}
\]

where \(m_{ij}\) and \(M_{ij}\) are the predicted and observed minutes, respectively, for the \(i\)th physician and \(j\)th patient.

Because fellowship training in pediatric emergency medicine (PEM) and the duration of practice can contribute to differences in resource use, we studied the effect of these 2 variables on OE ratios in all resource categories. For each resource, independently, we performed multiple linear regression analyses with 3 predictors (fellowship training, years in practice, and the 2-way interaction) and OE ratio as outcome.

Using Spearman correlation coefficient, we examined whether a physician’s relative use (OE ratio) in one resource category was correlated to that physician’s relative use in another category and whether relative use was correlated with the relative measures of ED quality (LOS and RR). All analyses were performed in SAS (SAS Institute Inc, Cary, NC) and R (R Development Core Team, Vienna, Austria) software.\(^{44,45}\)

### RESULTS

The following results reflect our analysis of 163,669 patient visits at ED1 attended by 36 attending physicians and 289,199 visits at ED2 attended by 45 attending physicians for the study period. The demographic and unique clinical characteristics are summarized in Table 1.

#### Emergency Department-Level Resource Use Patterns

Case-mix adjusted admission rates, laboratory tests ordered, imaging, IV therapy, LOS, and RR are shown in Table 2 expressed as mean, median, and range for physicians at each ED. Overall, 13.3% visits at ED1 and 12.1% at ED2 resulted in hospital admission; hospital admission rates varied almost 3-fold (6.3%–18.0%) among the 36 physicians at ED1 and almost 8-fold (2.5%–19.4%) among the 45 physicians at ED 2. Among patients not admitted, imaging rates showed a nearly 2-fold (10.9%–20.6%) variation at ED2. Variation among physicians was less pronounced for laboratory test ordering at both sites and for imaging at ED1. Head CT use was 1.9% at ED1 with a more than 2-fold variation among the 36 physicians (1.1%–2.5%) and 2.9% at ED 2 with a more than 5-fold variation among the 45 physicians (0.9%–4.8%). Approximately 7% of the patients received IV therapy: this varied 2-fold (4.9%–10.4%) among ED1 physicians and more than 3-fold (3.6%–11.4%) for ED2 physicians. The LOS distribution was strongly right-skewed: the mean and median adjusted LOS was longer at ED2, with more physician variation.

### Table 2. Case-Mix Adjusted Resource Use Characteristics of Physicians Practicing in 2 Pediatric EDs (2003–2006)

<table>
<thead>
<tr>
<th></th>
<th>ED1 (n = 36)</th>
<th>ED2 (n = 45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Visits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients seen/physician, mean, median (range)</td>
<td>4546, 4769 (1042–9296)</td>
<td>6426, 5661 (1008–17620)</td>
</tr>
<tr>
<td>Admissions, mean, median (range), † %</td>
<td>13.3, 13.8 (6.3–18.0)</td>
<td>12.1, 13.2 (2.5–19.4)</td>
</tr>
<tr>
<td>For patients sent home</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory tests ordered, mean, median, (range), † %</td>
<td>20.7, 20.6 (18.3–23.9)</td>
<td>22.3, 22.2 (17.6–26.2)</td>
</tr>
<tr>
<td>Imaging ordered, mean, median (range), ‡ %</td>
<td>14.4, 14.3 (10.3–16.4)</td>
<td>15.4, 15.8 (10.9–20.6)</td>
</tr>
<tr>
<td>Head CT, mean, median (range), %</td>
<td>1.9, 2.0 (1.1–2.5)</td>
<td>2.9, 3.3 (0.9–4.8)</td>
</tr>
<tr>
<td>IV therapy, mean, median (range), § %</td>
<td>7.1, 6.9 (4.9–10.4)</td>
<td>7.2, 7.5 (3.6–11.4)</td>
</tr>
<tr>
<td>LOS, mean, median (range), ‡ min</td>
<td>65, 65 (58–75)</td>
<td>75, 76 (61–96)</td>
</tr>
<tr>
<td>RR, mean, median (range)</td>
<td>3.8, 3.7 (3.1%–3.4%)</td>
<td>3.9, 3.9 (3.1%–5.0%)</td>
</tr>
</tbody>
</table>

Interquartile ranges were also computed (data not reported).

†Admission is defined as inpatient or observation status, operating room, intensive care, or transfer.

‡Laboratory tests include basic or comprehensive metabolic panel, complete blood count, and rapid strep test.

§Imaging includes chest radiography, abdominal radiography, and noncontrast head CT.

¶Intravenous therapy includes IV fluids and antibiotic therapy (ceftriaxone or cefotaxime).

‖Length of stay is defined as time in minutes from first examination by physician or midlevel provider to physician order for disposition home.

¶Return rate denotes return to ED within 72 hours of initial visit.
Physician-Level Resource Use Patterns for ED1

Individual physicians’ OE ratios for laboratory tests, imaging, IV therapy, and admissions at ED1 are shown in Figure 1, with OE ratios for each physician reported in ascending order. Each dot represents a single physician’s OE use ratio, with the vertical line showing the 95% CI. There were 2- to 3-fold differences in the minimum and maximum OE ratios for the 36 physicians at ED1: laboratory tests (0.7–1.62), imaging (0.62–1.81), and IV therapy (0.57–1.45), and more than 3-fold for head CT (0.47–1.80). For hospital admission, the OE ratio varied from 0.75 to 1.27.

Table 3 examines whether higher than expected resource use was associated with longer than expected ED LOS or higher/lower than expected RR. Physicians had a higher than expected LOS if they had higher OE ratios for laboratory tests ($r$ = 0.41; 95% CI, 0.09–0.65; $P$ < 0.05) and imaging ($r$ = 0.48; 95% CI, 0.17–0.69; $P$ < 0.01). Of note, there was no correlation between physicians’ relative use of any resource category and OE ratio for RR.

The correlation between physicians’ OE use ratios in different resource categories examines the strength of association between a physician’s use ratio for one resource and a physician’s use ratio for another resource. Physicians with higher than expected use of laboratory tests had higher than expected use of imaging ($r$ = 0.62; 95% CI, 0.36–0.78; $P$ < 0.001), higher than expected use of head CT ($r$ = 0.49; 95% CI, 0.19–0.70; $P$ < 0.01), and higher than expected use of IV therapy ($r$ = 0.51; 95% CI, 0.20–0.71; $P$ < 0.01).

No statistically significant correlation was found between PEM training status and any resource use or between years in practice and any resource use. Only the analysis for LOS showed a statistically significant difference ($P$ = 0.02). Our results indicate that as years in practice increase, physicians with no PEM training tend to have higher OE ratios for LOS (ie, their patients tend to have longer LOS than would be expected under the internal standard); the opposite trend seems to be true with physicians who have PEM training.

**DISCUSSION**

Our study underscores wide variation in case-mix adjusted resource use by physicians in the pediatric ED, in both academic and private settings. We evaluated this variation by studying the range of overall adjusted utilization and by calculating OE ratios; both methods demonstrate up to 8-fold variation in resource use in the categories studied. We believe ours is the first study of its kind to demonstrate case-mix adjusted variation in resource use by attending physicians within a single specialty (PEM), in a single practice setting (pediatric ED), across a wide array of presentations and conditions typical for an ED.

The 2001 Institute of Medicine report, *Crossing the Quality Chasm*, notes the importance of quality improvement efforts within microsystems that form a basic unit of care delivery. The report recommends measurement of unscientific variation in health care delivery, moving from administrative and service processes to clinical issues. The Agency for Healthcare Research and Quality's *Guide to Clinical Effectiveness* provides practical information about clinical effectiveness research, including guidelines for conducting systematic reviews and meta-analyses, and evidence-based practice guidelines for a range of clinical conditions. These resources are valuable tools for clinicians and researchers alike, offering a comprehensive understanding of the latest research and best practices in the field of emergency medicine.
and Quality also recently emphasized the importance of comparative effectiveness research by selecting quality measures in areas with high-performance variation. Furthermore, recent discussions on quality measurement have suggested measuring quality at the level of individual practitioners, for example, the Ongoing Professional Practice Evaluation standard.

In our study, use of IV therapy had a 2- to 3-fold variation even after severity adjustment. The decision to treat with IV fluids/antibiotics can be discretionary in the ambulatory setting: IV fluids are sometimes ordered simply because the patient has an IV; the IV is often placed simply because a patient is getting venipuncture for laboratory tests. This type of decision making can result in a sequential cascade of resource use that is potentially avoidable.

There has been a recent upward trend in the use of diagnostic imaging in general and CT in particular, in both children and adults. A recent study showed increasing use of CT scans for children in the pediatric ED. Our study shows a 2- to 5-fold variation between physicians in overall adjusted use of head CT. Cost issues notwithstanding, unnecessary CT can result in significant radiation exposure. However, there are limited guidelines on when to perform (or when not to perform) CT in the ED setting. Ready availability of CT scanners in most hospital EDs, along with financial incentives and defensive practice patterns, can result in overuse of such imaging modalities. An important first step in optimizing use would be to study physician patterns of CT use to determine if large variation exists. Provision of comparative data such as those presented in this study can be a powerful motivator for changing clinician behavior.

There was a 3- to 8-fold difference in case-mix adjusted hospital admission rates. The Dartmouth Atlas project shows wide variation in rates of hospital admission among Medicare patients in their last 2 years of life, leading authors to question what the appropriate admission rate is. Goodman et al found that the likelihood of hospitalization for pediatric medical conditions for which outpatient alternatives are available is influenced by the supply and the character of medical care. According to a recent report, approximately 50% of all pediatric hospital admissions begin in the ED. Although there are some definitive indications for hospital admission, for many conditions, the decision to hospitalize versus outpatient therapy can be discretionary. Arguably, a hospital admission is perhaps the most resource intensive decision an ED physician makes and one that can lead to another cascade of resource use opportunities.

Our study attempts to show how resource use correlates with ED efficiency by evaluating its effect on LOS. Higher than expected resource use resulted in longer than expected LOS particularly for laboratory tests and imaging. Although this is not unexpected, physicians tend to argue that their higher LOS resulted from sicker patients. We found that even after adjusting for severity of illness and case-mix, patients of physicians who typically do more testing/treatment have longer ED LOS. Reducing this unnecessary variation in testing and treatment has the potential for reducing ED LOS and ED crowding.

We also studied whether lower test/treatment use would result in an increase in unintended consequences. Because mortality is very rare in pediatrics, we used a 72-hour return to the ED as a marker for potentially unmet patient care needs that necessitated return to care. We found no relationship between RR and test/treatment use in any category; physicians with high testing and treatment use have RRs equivalent to those who are less aggressive with testing and treatment, suggesting that doing more might not be adding diagnostic or prognostic value to the care in the ED.

### Trends Across Utilization in Different Resource Categories

Physicians with high utilization in one resource category tended to have high use in other categories as well. Some previous studies have evaluated the effect of practice style on ordering of tests and therapies. Although our study was not intended to study practice style or the factors associated with it, we do show that this may have a role in contributing to variation in resource use.

There is a general perception that fellowship training and increasing years in practice influence resource use. This can potentially go in either direction: increasing testing or treatment because of past adverse outcomes or decreasing it because of experience and comfort level. Although our study was not designed to specifically study such differences, we did find that fellowship training and duration of practice did not correlate with resource use in any category.

Development of meaningful clinical quality metrics is an important quality improvement agenda. Gutman et al evaluated a systematic approach for developing clinical performance indicators for the ED care of children; many of their suggested indicators include measures of diagnostic test use, particularly radiography. Despite the lack of benchmarks for the right rate, they recommend that EDs track measures of resource use to identify areas with an opportunity for improvement. A recent study by Knapp et al on ED care of children with asthma, bronchiolitis, and croup found underuse of proven therapies and overuse of diagnostic studies and treatments. Using case-mix adjusted resource use as an ED quality metric has the potential to reduce variation and optimize testing, treatment, and LOS in the ED.

### STRENGTHS AND LIMITATIONS

The study uses retrospective analysis of data extracted from an existing EMR database. However, evaluation of variation in practice is best studied using a retrospective design to avoid any Hawthorne effect.

We evaluated return within 72 hours only to the same ED. It is possible that a few patients may have presented to an outside health care provider within 72 hours; however, in our comparison of physicians, all were likely to be affected similarly by factors related to patient return visits.

Laboratory tests, imaging, and IV therapy were all analyzed as dichotomous outcomes, although each is composed of multiple possible resource uses. For example, a patient receiving all

---

**TABLE 3. Correlation Between OE Ratios for Resource Use (Laboratory Tests, Imaging, Head CT, and IV Therapy) and ED Quality Metrics (LOS and RR) for Discharged Patients in ED**

<table>
<thead>
<tr>
<th>Resource Category</th>
<th>LOS, r (95% CI)</th>
<th>RR, r (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory tests</td>
<td>0.41 (0.09 to 0.65)</td>
<td>-0.16 (-0.46 to 0.18)</td>
</tr>
<tr>
<td>Imaging</td>
<td>0.48 (0.17 to 0.69)</td>
<td>0.001 (-0.33 to 0.33)</td>
</tr>
<tr>
<td>Head CT</td>
<td>0.27 (-0.07 to 0.55)</td>
<td>-0.04 (-0.36 to 0.29)</td>
</tr>
<tr>
<td>IV therapy</td>
<td>0.14 (-0.20 to 0.45)</td>
<td>0.02 (-0.31 to 0.35)</td>
</tr>
</tbody>
</table>

*Length of stay denotes time from first examination by physician or midlevel provider to physician order for disposition home.

†Rate of return to ED within 72 hours of initial visit.

‡Laboratory tests include basic or comprehensive metabolic panel, complete blood count, and rapid strep test.

§Imaging includes chest radiography, abdominal radiography, and/or noncontrast head CT.

‖Intravenous therapy includes IV fluids and/or antibiotic therapy (ceftriaxone or cefotaxime).
types of imaging was counted equivalently to a patient receiving just one. However, our finding that physicians who overused in one category also ranked high in another suggests that if test ordering within a category was counted cumulatively, it may have resulted in an even wider variation in resource use.

We acknowledge 2 caveats in our post hoc analyses of physician training and duration of practice. First, the absence of a statistical significance with resource use may be owing to lack of statistical power with only 36 physicians. Second, the significant trend noted for LOS was heavily influenced by 1 data point. If that point is removed, no significant trend is found.

Finally, this analysis is limited to pediatric ED patients in one city and may not be generalizable to other EDs.

CONCLUSIONS

Significant variation exists in physician use of common resources in the pediatric ED, even after risk-adjustment for severity of illness and case mix. We demonstrate its presence in both academic and nonacademic settings. Physicians using more resources than expected have longer than expected lengths of stay in the ED but do not have lower than expected rates of return to the ED. Physicians with higher than expected use in one resource category also have higher than expected resource use in other categories. Studies evaluating practice variation, followed by efforts to impact practice patterns of individual practitioners, can form an important addition to other quality improvement measures.

REFERENCES


46. Crossing the Quality Chasm: A New Health System for the 21st Century Institute of Medicine.


The unbearable lightness of diagnostic testing: time to contain inappropriate test ordering

A Schattner

ABSTRACT

The numbers of tests, imaging and diagnostic procedures ordered by doctors in primary care and hospital settings alike are rising. According to established guidelines, many of these tests may be unnecessary. Various non-clinical factors often underlie test ordering, and multiple prevalent problems associated with tests and their interpretation are often under-appreciated. These include the significant potential for cascades of further testing and patient harm. Multifaceted strategies may improve test-ordering behaviour. Brief reconsideration of several points by the clinician before ordering the test is advisable.

Doctors are all very fond of tests. Tests are very easy to order, they cost (us) nothing, the patient is impressed and pacified, and, until the test is scheduled, performed and interpreted, we win that most desirable of all commodities—time. Even if we are not too sure of what we are looking for, there is always the chance that the test will come up with something—at the very least, something, that can be followed-up by another test.

With the growing number and escalating costs of modern tests, “the unbearable lightness of testing” constitutes an increasing burden on patients, doctors and society. Clinicians tend to disregard a number of problems associated with test ordering and with the interpretation of their results. This article discusses a few of those problems, not because they are new or unfamiliar, but because they are highly prevalent and rarely considered together, despite their significant impact on the costs and quality of healthcare.

THE TEST IS REDUNDANT

Test ordering is far from being a pure biomedical decision founded on a solid scientific rationale. The patient–doctor interaction is also a social interaction, and decisions about tests may be influenced by the doctor’s personality, the doctor’s emotional state, numerous patient characteristics (age, gender, look, status, insurance coverage), perceived patient pressure and other variables. When a decision to test is closely linked to the doctor’s own remuneration, judgment may well be biased in favour of testing. This may be either due to the recently introduced fee-for-service system or because the doctor controls the testing apparatus. Moreover, 60% of doctors may order more tests than medically indicated for defensive reasons. This pattern adds to poor accountability and the extreme ease of ordering almost anything—computer-based test ordering is now practically effortless. The patient’s problem is postponed, and the doctor immediately gains time. The result is an appalling volume of redundant testing in both hospital settings and ambulatory care. It has been found that simple enough interventions are able to reduce routine laboratory tests and chest radiography by more than half with no worsening of clinical outcomes. Frequent repetition and inappropriate ordering to relieve stress from uncertainty are common and could be greatly improved. A bias exists towards new, prominent and prestigious diagnostic technologies. Studies expose their accelerated adoption years before any solid evidence of efficacy. Such evidence may never materialise. Increased diagnostic testing is closely followed by more diagnostic procedures and frequent therapeutic interventions. However, population survival benefits are uncertain and the merits of such hyperactivity remain controversial.

THE TEST IS ATTEMPTED BUT FAILS

Numerous factors may lead to a technical failure of the test. They include faulty patient preparation, poor cooperation and incompatibility between the patient’s characteristics and the test. Thus, a lot of anxiety and frustration may be experienced, much time may be lost to no avail, and, sometimes, frank adverse events may result. Although the actual proportion of failed diagnostic procedures has never been studied to my knowledge, experience shows that they are not negligible.

THE TEST IS ORDERED OR INTERPRETED OUT OF ITS CLINICAL CONTEXT

The patient’s history must first be elicited and the patient examined before the necessity of a test, its selection and interpretation can be determined. The clinical information is crucial for decisions. Thus, a 50-year-old man with chest pain may be safely sent home if the history, examination and ECG are normal. However, he may need: a workup to rule out pulmonary embolism if the pain is pleuritic; echocardiography if an aortic murmur is detected; a stress thallium test if multiple risk factors are found. Proper estimation of the pretest probability (PTP) and familiarity with the test’s sensitivity, specificity and likelihood ratios may prevent major errors. Specific risk factors, typical clinical presentation and lack of an alternative explanation may increase the PTP, and the prevalence of the suspected condition in a similar population should be considered. The higher the PTP, the more sensitive the selected test needs to be to rule out a suspected diagnosis. For example, a normal d-dimer (sensitivity 96%) may exclude pulmonary embolism unless the PTP is high. Despite current extensive sophisticated tests,
diagnostic mistakes persist. Surprisingly perhaps, the history and examination retain their prime value as an essential gate to diagnostic testing. Recent studies on patients with prototype enigmatic diagnoses, such as fever of unknown origin or syncope, reaffirm this conclusion. However, doctors today may perform but a perfunctory history and examination, relying on type enigmatic diagnoses, such as fever of unknown origin or syncope, to solve the patient’s problem.

**Box 1: Questions to ask before ordering tests**

**History and examination**
Did I take a good history and examine the patient? Do I have all the clinical facts?

**Question**
What exactly is the clinical question that the test addresses?

**Test characteristics**
Is the test suited for this purpose (bearing in mind pretest probability and the test’s sensitivity/specificity)?

**Alternative**
Is there a viable alternative option? For example, just watching the patient or choosing a less risky test.

**Repetition**
Did I check that the test does not unnecessarily repeat what was already done? Would it really add new information?

**Potential for adverse outcome**
Is there a danger to the patient from the test? Is the clinical benefit worth the potential risk?

**Informed consent**
Did the patient consent to the test after understanding what is being planned?

**Preparation**
Is patient preparation optimal?

**Communication with the tester**
Was the person performing the test provided with the necessary clinical information?

**Follow-up**
How and when will the test results be retrieved, communicated to the patient and acted upon?

THE ABNORMAL RESULT IS FALSE-POSITIVE

Even 95% specific tests will probably yield one false-positive result per 20 tests \((1-(0.95)^{20})\) when multiple determinations are requested for healthy people. Most tests have a lower specificity, and, when performed in older or sick people, false-positive results abound and permeate every method of testing. Cancer screening is perhaps the most susceptible: of 502 women screened for ovarian cancer by ultrasound scan, 21% required further scanning for suspicious results and 7% were referred for surgery. Not one of the patients was found to have cancer. Of 10 000 women who start mammography screening at the age of 40 years, half might receive at least one false-positive result over a decade of annual screening. In addition, a small percentage of results are liable to a laboratory mistake (<0.5%). Significant fluctuations in level or titre may also be encountered.

These considerations need to be made before testing. It is helpful to consider current evidence-based guidelines, the risk/benefit ratio and the patient’s preferences. Inappropriate application or interpretation of tests can rob people of their perceived health and be detrimental, costly and time consuming.

THE ABNORMAL RESULT IS ACCURATE BUT NON-SPECIFIC OR OF DOUBTFUL RELEVANCE

When practically hundreds of measurements and pictures are available after each test, it is hardly surprising that the number of pathological findings has escalated. However, their relevance to the patient’s symptoms and future impact may be small. The dubious significance of the findings may pose serious questions about the wisdom of ordering the test in the first place. The fact that about one in every two healthy adults who underwent cardiac multidetector CT had incidental findings, and 23% were recommended for further investigations indicates the magnitude of the problem. The use of CT or MRI for patients with low back pain raises questions, when even healthy asymptomatic people may have serious abnormalities of the spine upon imaging. Endocrine incidentalomas (thyroid, adrenal, pituitary) are commonly found during imaging, initiating seemingly endless cascades of further testing, follow-up and distress. The large majority are clinically insignificant. Tests often expose prevalent pathologies (eg, cholelithiasis, haemorrhoids, Helicobacter pylori), which may be incidental and unrelated to the patient’s symptoms. Clinical judgement remains crucial.

In addition, doctors may mistakenly attribute diagnostic qualities to tests with a limited specificity. Thus, most patients with serum autoantibodies (eg, antinuclear autoantibodies, rheumatoid factors) do not have an autoimmune disease and may even be healthy. Most patients with positive D-dimer do not have venous thromboembolism, and other examples abound. Thus, the decision rests on careful clinical judgement and on the recognition of the limitations of tests, as well as their power.
Faculty matters

Box 2: Interventions that may reduce inappropriate ordering of laboratory tests, imaging and diagnostic procedures

Educational
- Improved patient–doctor communication
- Time management skills
- Better problem solving
- Better informed test selection and understanding of test characteristics
- Dealing with uncertainty in medicine and developing tolerance
- Economic aspects of test ordering
- Dissemination and discussion of evidence-based guidelines
- Continuous quality improvement discussions in small groups

Institutional
- Tests unbundled to specific requests
- Supervision of test ordering by senior doctors
- Restricted access to tests, according to level of training/seniority
- Displayed charges concurrent with test ordering
- Computerised reminders when tests are repeated unnecessarily
- Regular feedback/audit on the use of tests, compared with peers
- Economic incentives rewarding savings resulting from appropriate use of tests

Combinations
Multifaceted strategies combining several modalities were associated with significantly improved test-ordering performance\(^1\)\(^2\)\(^3\)

THE TEST IS ABNORMAL AND SIGNIFICANT BUT DISREGARDED
Current systems in healthcare do not reliably ensure that test results are monitored, received and acted upon by the ordering doctor.\(^4\)\(^5\) A third of doctors failed to notify their patients when test results were abnormal.\(^6\)\(^7\) Even results disclosing cancer and infectious diseases were not acted upon. This was common in both hospitals and clinics and linked to serious harm in some cases.

THE TEST IS ASSOCIATED WITH PATIENT HARM
The uncertainty while waiting for a test, and then for its results, may be associated with considerable patient anxiety and distress.\(^8\) The wait may be quite prolonged in cases referred for repeat or additional tests. Distress may persist even after a normal result is obtained.\(^9\) The test itself may be uncomfortable or painful to the patient, but test-associated harm is the more substantial threat. The increasing age, comorbidity and multidrug use of patients today substantially enhances their risk of developing complications after tests. Contrast-induced nephropathy is one well-studied example. A downward spiral of medical activities may develop as one test quickly and “inevitably” leads to another, and procedures and operations follow in uncontrollable, unstoppable fashion.\(^10\) Such “cascades” may culminate in severe patient harm and even death.\(^11\) Moreover, many imaging tests, especially when performed in childhood or repeatedly, are also a source of radiation exposure. The radiation dose from one CT scan is typically comparable to that of several hundred x-ray examinations. The sharp increase in its use\(^1\) can probably be associated with a small but significant increase in the risk of cancer.\(^12\)\(^13\) Viewed in this light, the estimation that one-third of all CT scans could be avoided\(^14\) is disturbing. When the decision to test may easily become a decision to treat, and tests and treatment options are far from benign, more attention should be given to the clinical details and alternatives before tests are ordered.

THE TEST CAUSES INDIRECT HARM BY ITS OPPORTUNITY COSTS
Doctors today must add the economic impact of their actions to their considerations. Besides squandering resources, overburdening the system with tests that are not really indicated may block access of other more needy patients to the same test, potentially delaying important diagnoses and adversely affecting healthcare.

COMMENT
These 10 considerations are but a reminder that, far from being simple, straightforward and always accurate, tests are the first step and are liable to many potential interferences before they can be used to the advantage of the patient. Admittedly, none of the problems cited may occur. However, several such difficulties will often combine, significantly compromising patient care. It is our duty to remember these possibilities when ordering and interpreting a test.

As a pre-emptive measure, brief consideration of several key questions (box 1) before a test is actually ordered may help the clinician to achieve more optimal use of tests.

The recent debate on the rapidly proliferating use of CT angiograms to visualise the coronary arteries (\textit{New York Times}, 29 June 2008) exemplifies many of the problems discussed. Unfortunately, in this typical example, financial incentives seem to dwarf clinical sense. Their use is spreading despite the meagre supporting evidence, substantial radiation exposure and high risk of further poorly warranted testing.

In conclusion, the power of the timely use of the correct test for the right patient after adequate preparation by a competent clinician is indisputable. However, barely discriminated polystaining should be condemned, monitored and held in check. Doctors—throughout medical school, residency programmes and beyond—need better education on problem solving, test selection and operating characteristics, dealing with uncertainty, improved communication with patients, and economic aspects of test ordering. Several other interventions at different educational and institutional levels have been tried with varying efficacy (box 2). The accumulating evidence supports multifaceted approaches as the most efficacious in bringing about a change.\(^2\)\(^3\) Until then, we could all benefit from a more discerning and balanced approach to diagnostic testing.

Competing interests: None.
REFERENCES


BMJ Masterclasses

BMJ Masterclasses are educational meetings designed specifically to meet the learning needs of doctors. They help doctors keep up to date with the latest evidence and recent guidelines in major clinical areas, enabling them to use the latest evidence to make better decisions. The latest evidence, recent guidelines and best practice are delivered in an interactive and informative manner by leading experts. The speakers are specifically chosen as highly-skilled communicators who can authoritatively enthuse the audience and interpret the latest research and guidelines into practical tips for busy doctors. BMJ Masterclasses have produced a huge hit with clinicians, with many saying they have influenced their clinical practice. http://masterclasses.bmj.com/
The unbearable lightness of diagnostic testing: time to contain inappropriate test ordering

A Schattner

Postgrad Med J 2008 84: 618-621
doi: 10.1136/pgmj.2008.072579

Updated information and services can be found at:
http://pmj.bmj.com/content/84/998/618.full.html

These include:

References
This article cites 43 articles, 12 of which can be accessed free at:
http://pmj.bmj.com/content/84/998/618.full.html#ref-list-1

Article cited in:
http://pmj.bmj.com/content/84/998/618.full.html#related-urls

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

General practice / family medicine (95 articles)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/