Abstract: Background: A pediatric abdominal examination requires a delicate balance of gentleness and thoroughness, which prevents undue discomfort, but allows the pediatric physician to arrive at the correct diagnosis. We hypothesize that there is a significant difference in abdominal examinations between a general surgery resident (GS) and an emergency department resident (ED).

Methods: The documented abdominal examinations for 100 consecutive children evaluated for possible appendicitis by both an ED and GS resident at a children's hospital were compared.

Results: The ED examiners tested for and documented rebound tenderness \( (p = 0.0072) \) and bowel sounds \( (p < 0.0001) \) significantly more than the GS examiner. The GS examiners tested for and documented the presence of a soft abdomen, guarding, peritoneal signs, psoas, Rovsing's, and obturator sign significantly more than the ED examiner (all with \( p \) value of \( \leq 0.001 \)). When the ED examiner recorded guarding, only 2 of 19 specified whether guarding was voluntary or involuntary, compared to 21 of 36 GS examiners \( (p = 0.0006) \).

Conclusion: An excessive degree of variability and disagreement exists in pediatric abdominal examination, which may contribute to unnecessary diagnostic testing and associated radiation exposure, as well as diagnostic delay and potential patient morbidity.
Who’s guarding whom? The changing of the guard: A study of pediatric abdominal examinations

Robert J Doiron, BS¹, Charles W Hartin, Jr., MD², Jeffery M Jordan, MD, PhD³, Doruk E Ozgediz, MD, MSc¹,² Philip L Glick, MD, MBA¹,²,³

Affiliations

¹ State University of New York at Buffalo, School of Medicine, 40 BEB, 3435 Main Street, Buffalo, NY 14214

² Department of Pediatric Surgical Services, Women & Children's Hospital of Buffalo, 219 Bryant Street, Buffalo, NY 14222

³ Department of Surgery, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, 100 High Street, Buffalo, NY 14203

Correspondence and Reprints to:
Philip L Glick, MD, MBA
Department of Pediatric Surgery
Women & Children’s Hospital of Buffalo
219 Bryant Street
Buffalo, NY 14222
Phone: 716-878-7449
Fax: 716-878-7998
Email: glicklab@aol.com
Abstract

**Background:** A pediatric abdominal examination requires a delicate balance of gentleness and thoroughness, which prevents undue discomfort, but allows the pediatric physician to arrive at the correct diagnosis. We hypothesize that there is a significant difference in abdominal examinations between a general surgery resident (GS) and an emergency department resident (ED).

**Methods:** The documented abdominal examinations for 100 consecutive children evaluated for possible appendicitis by both an ED and GS resident at a children’s hospital were compared.

**Results:** The ED examiners tested for and documented rebound tenderness \((p = 0.0072)\) and bowel sounds \((p < 0.0001)\) significantly more than the GS examiner. The GS examiners tested for and documented the presence of a soft abdomen, guarding, peritoneal signs, psoas, Rovsing’s, and obturator sign significantly more than the ED examiner (all with \(p\) value of \(\leq 0.001\)). When the ED examiner recorded guarding, only 2 of 19 specified whether guarding was voluntary or involuntary, compared to 21 of 36 GS examiners \((p = 0.0006)\).

**Conclusion:** An excessive degree of variability and disagreement exists in pediatric abdominal examination, which may contribute to unnecessary diagnostic testing and associated radiation exposure, as well as diagnostic delay and potential patient morbidity.
Introduction

A good history and physical examination have been the backbone of medicine and are essential to guide diagnostic tests and arrive at the patient’s correct diagnosis. Medical school and residency curricula are designed to provide a physician with the tools to perform and document standardized physical examination components. Physical examination skills are so important that the National Board of Medical Examiners (NBME) implemented clinical skills (CS) testing for all United States (US) medical graduates finishing in the year 2005 or later [1]. As part of the United States Medical Licensing Examination (USMLE) Step 2, this examination ensures that a minimum standard for all licensed physicians regardless of the site of their training [2].

Recent studies have demonstrated that the accuracy of diagnosis in adult patients is consistent between different specialist examiners [3,4]. However, a child’s examination requires an expanded skill set, and the abdominal examination is a critical component of evaluating abdominal pain in children [5]. When done improperly in a child, the physical examination can cause unnecessary pain and anxiety and demonstrate a confusing constellation of findings. For example, a child with abdominal complaints may have voluntary guarding as a result of multiple examinations, and the continued elicitation of “rebound tenderness” by an examiner in a scared child is likely a painful and unnecessary misadventure. However, a thorough, but gentile abdominal examination in a child with early appendicitis may demonstrate a soft abdomen with mild involuntary guarding at McBurney’s point leading to a correct diagnosis and avoiding unnecessary imaging studies.
We hypothesize that there are significant differences in the examination and interpretation of a child’s abdomen between an emergency department (ED) resident and a general surgery (GS) resident that may prolong workup and delay definitive care.

Methods

We conducted an Institutional Review Board approved retrospective chart review at a pediatric tertiary emergency department, which handles approximately 45,000 patients per year. The records were retrospectively reviewed for 100 consecutive children ages 2 to 18 presenting with abdominal pain and examined by both ED and GS residents. ED residents included post-graduate year (PGY) 1 – 3 in the State University of New York at Buffalo’s accredited emergency medicine or pediatric residency programs. GS resident examiners were all PGY-3 and part of the university’s accredited general surgery program. All residents passed the USMLE Step 2 examination as a requirement for admission to the university graduate medical education programs.

In all cases, surgery was consulted for suspected appendicitis between the dates of January 2010 through December 2010. No patients were excluded. During this time period, both examiners recorded their physical examination findings as written free hand text without prompts for different components of the examination. The documented critical components of an abdominal examination were recorded for each examiner and included the following: soft abdomen, guarding (voluntary versus involuntary), rebound tenderness, masses, distention, peritoneal signs, bowel sounds, and tenderness to palpation. Psoas, obturator, or Rovsing’s sign were also recorded. The findings between
the groups were evaluated for statistical differences using the Fisher exact test with a $p$
value of 0.05 being defined as significant.

**Results:**

Of the 100 patients reviewed, the average age was 11 with 63% males. Table 1
demonstrates the number of times each group documented physical examination
components (either positive or negative). There were significant differences between the
groups in documenting each component. The GS examiners documented whether the
patient’s abdomen was soft, had guarding, or had peritoneal signs (or peritonitis)
significantly more than the ED examiner. In addition, there was a statistically significant
increase in documentation of psoas, Rovsing’s, or obturator signs by GS residents.
However, the ED examiner documented the status of bowel sounds and presence or
absence of rebound tenderness significantly more than the GS examiner.

A more concerning difference between the groups was the discrepancy in the
interpretation of physical examination findings. Table 2 demonstrates their actual
responses in either yes (positive) or no (negative) for the given physical examination
component of the same 100 children. There was little to no agreement among the groups
for most of the physical examination components that were documented as abnormal.
The highest agreement between the ED and GS examiners was found when documenting
non-tenderness (28% agreement) or non-specified guarding (34% agreement).

When the ED examiner indicated that there was positive guarding, only 2 of 19
differentiated voluntary and involuntary guarding. The GS examiner documented
voluntary or involuntary guarding in 21 of the 36 children. This difference between groups is also statistically significant ($p = 0.0006$).

**Discussion**

An inadequate abdominal examination has been shown to increase imaging and delay diagnosis, which can lead to increased patient morbidity [6-12]. In addition to diagnostic delay, unnecessary imaging resulting from inaccurate physical examination increases the cost of care and radiation exposure to the child [6-13]. Radiation exposure in children is especially important as they are more sensitive and 1 per 1000 pediatric computed tomography (CT) scans have been estimated to cause a potentially fatal malignancy [14-16]. CT scanning is often employed to make the diagnosis of appendicitis [8-12]; however, some studies have indicated that CT scans are no better than a pediatric surgeon’s performance of an adequate history and physical examination [7].

Since physical examination can be as effective as imaging to diagnose appendicitis, accurate and reproducible findings are imperative among the diverse practitioners who evaluate pediatric abdominal pain. One recent study compared the accuracy of diagnosis of appendicitis by physical examination between emergency medicine residents and surgical residents [3]. They concluded that because of a close correlation between the ED and GS resident examinations that early surgical consultation provided no benefit. This study conducted in Korea excluded patients younger than 15. Our study did not reproduce the same results with a younger population.
An US study in 2005, before the standardized USMLE CS examination, reported the interexaminer reliability of 68 pediatric abdominal examinations between pediatric surgeons and emergency department personnel [17]. In this study, they determined that no component of the physical examination was reliable between groups. With our study demonstrating similar results for 100 children, it would appear that standardizing medical education curricula and additional physician testing for clinical skills has not improved these discrepancies. Other studies have had variable results but none to our knowledge have compared similar trainees in the pediatric abdominal examination since implementation of the new USMLE testing standard [4].

Documentation of different components of the physical examination likely demonstrates what the physician determines to be important and encompasses the culture of training in each specialty. In our study, the ED examiner documented bowel sounds and rebound tenderness significantly more than the GS examiner. This indicates that the GS examiner may not be using bowel sounds in the (surgical) decision tree for an acute abdomen as frequently as the ED examiner, who is also examining for medical causes of an acute abdomen. Some surgical faculty discourages the elicitation of rebound tenderness in children and believes it causes undue pain without changing management [5]. The GS examiner more commonly documented the components of soft abdomen, guarding, and peritoneal signs over the ED examiner. A GS resident’s training likely stresses these elements in decision making. Although not accessed by this study, rectal examination by either examiner was rarely performed but is often questioned by plaintiff’s attorneys in cases of missed appendicitis [18]. The relative diagnostic value of
each component of the abdominal examination is beyond the scope of this paper but does highlight the significant disparities.

Guarding is an important component of the pediatric abdominal examination, and the difference between voluntary and involuntary guarding should be established due to its clinical implications in the acute abdomen [19-21]. Involuntary guarding can suggest severe peritoneal irritation while voluntary guarding may be of no significance. In our study, the ED examiner specified guarding only 11% of the time compared to 58% for the GS examiner. Again, we believe prior training of each examiner contributed to the observed differences.

This study was performed at one institution and may limit generalizability to other institutions. No correlation was made between physical examination findings between the groups and any imaging, surgical findings, or pathological diagnosis. Although, our study was retrospective it may help eliminate bias via the Hawthorne effect of a prospective study. The Hawthorne effect may cause the different groups of examiners to alter their normal practice because they are aware that their findings are under review as part of a study. The kappa statistic has sometimes been employed to compare interrater reliability among physical examinations; however, our agreement between groups was so poor that most of the values in the kappa statistic would be “0” and be of little value to report [22]. Another limitation is the differences of training levels between the examination groups. Prior studies have indicated that more junior trainees document physical examination findings inconsistent with that of senior level trainees and attendings [4,23]. However, these studies were also prior to the USMLE CS. As noted above, all examiners in our study have passed the USMLE CS.
The GS examiner in our study was typically consulted after the ED examination and usually also after imaging. This may have biased the GS examiner and influenced the documented physical examination findings. In addition, the time period between the ED and GS examination was not assessed by this study. The patient may have received pain medication or antibiotics between examinations, and this could have also affected physical examination findings. Some degree of inconsistency between the groups would be expected based on the above; however, we believe the level of inconsistency between the groups is excessive. The authors believe that education or perhaps decision making protocols should be aimed at reducing the discrepancies in physical examination findings and documentation between the groups. Multidisciplinary training could include the involvement of surgical and pediatric faculty in a medical school’s physical diagnosis classes or sessions tailored to emergency department staff [24].

In conclusion, our study demonstrates that in an era of performance based improvements, continuing medical education (CME), and increased testing (USMLE CS) and standardization of curricula, we have lost focus on some of the basic principles of diagnosing our patients [25]. The loss of physical examination skills likely leads to a delay in diagnosis, increase costs, and excessive diagnostic testing. We believe that the inconsistency in both documentation and interpretation of the critical components of a physical examination suggests a need for curriculum changes and evaluation of the effectiveness of standardized testing of clinical skills, perhaps even at the level of post-graduate training in various specialties.
Legends:

Table 1 Differences in documentation between the groups of abdominal examiners.

Table 2 Interpretation and agreement between the groups of abdominal examiners.

References:


Table 1 Differences in documentation between the groups of abdominal examiners.

<table>
<thead>
<tr>
<th>Documentation of Component</th>
<th>Emergency Department Examiner</th>
<th>Surgical Examiner</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft Abdomen</td>
<td>67</td>
<td>94</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Guarding</td>
<td>40</td>
<td>73</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Rebound tenderness</td>
<td>44</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>Abdominal masses</td>
<td>9</td>
<td>13</td>
<td>0.499</td>
</tr>
<tr>
<td>Abdominal distention</td>
<td>100</td>
<td>99</td>
<td>1.00</td>
</tr>
<tr>
<td>Peritoneal signs</td>
<td>16</td>
<td>39</td>
<td>0.0004</td>
</tr>
<tr>
<td>Normal bowel sounds</td>
<td>56</td>
<td>17</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Tenderness to palpation</td>
<td>95</td>
<td>97</td>
<td>0.721</td>
</tr>
<tr>
<td>(+) Psoas</td>
<td>20</td>
<td>42</td>
<td>0.0012</td>
</tr>
<tr>
<td>(+) Rovsing’s</td>
<td>13</td>
<td>39</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>(+) Obturator</td>
<td>7</td>
<td>33</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>Emergency Department Examiner</td>
<td>Surgical Examiner</td>
<td>Group comparisons</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>(Yes)</td>
<td>(No)</td>
<td>(Yes)</td>
</tr>
<tr>
<td>Soft Abdomen</td>
<td>62</td>
<td>5</td>
<td>93</td>
</tr>
<tr>
<td>Guarding</td>
<td>19</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>Rebound tenderness</td>
<td>15</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Abdominal masses</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Abdominal distention</td>
<td>8</td>
<td>92</td>
<td>14</td>
</tr>
<tr>
<td>Peritoneal signs</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Normal bowel sounds</td>
<td>44</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Tenderness to palpation</td>
<td>87</td>
<td>8</td>
<td>82</td>
</tr>
<tr>
<td>(+) Psoas</td>
<td>6</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>(+) Rovsing’s</td>
<td>4</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>(+) Obturator</td>
<td>2</td>
<td>5</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 2** Interpretation and agreement between the groups of abdominal examiners.
January 5th, 2012

Editors-in-Chief SURGERY

Andrew L. Warshaw, M.D.
W. Gerald Austen Distinguished Professor Surgery
Harvard Medical School
Surgeon-in-Chief Emeritus
Massachusetts General Hospital
Boston, MA 02114 USA

Michael G. Sarr, M.D.
Department of Surgery
Mayo Clinic and Mayo Foundation
Rochester, MN 55902 USA

Dear Drs. Warshaw and Sarr,

Please consider our original report manuscript entitled "Who’s guarding whom? The changing of the guard: A study of pediatric abdominal examinations" for publication in SURGERY.

This material has not been previously published or submitted and will not be sent to another journal until a decision is made by SURGERY. This is an SUS paper and will be presented at the 7th Annual Academic Surgical Congress in Las Vegas in February 2012.

No personal conflicts of interest exist and no sources of funding were secured for this study.

We believe this paper demonstrates the need for additional education in performance of the pediatric abdominal examination.

Thank you very much for the review and consideration of our manuscript for your journal and please feel free to contact us if we can be of any assistance.

Sincerely,

Charles (Chip) W Hartin, Jr, MD
Philip L Glick, MD, MBA