Patterns of Misrepresentation of Clinical Findings on Patient Notes During the COMLEX-USA Level 2-PE

Jeanne M. Sandella, DO; Larissa A. Smith, PhD; Laurie A. Gallagher, DO; and Erik E. Langenau, DO

Context: A health care team depends on the accurate documentation of the patient-physician encounter, be it written or electronic. If documentation is inaccurate or incomplete, patient care may be adversely affected. Previous studies have identified factors that influence documentation errors, such as fatigue, carelessness, and being overworked. More research, however, is needed into the patterns of errors and, by extension, these patterns’ potential effect on patient care. Insights about these areas would be valuable to practicing physicians, as well as to medical educators, who could incorporate such insights into the training of medical students.

Objective: To identify potential patterns of commissive documentation errors on SOAP notes during the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE); specifically, to calculate the frequency with which specific items are misrepresented in SOAP (subjective, objective, assessment, plan) notes and to explore patterns of misrepresentation in relation to the documentation of history taking and physical examination.

Methods: Among the 12,510 candidates tested between July 2007 and June 2010, there were 24 candidates who failed the COMLEX-USA Level 2-PE because they misrepresented clinical findings on SOAP notes. These errors of commission (over-documentation) were categorized as errors in either history-taking or physical examination, and then subcategorized and analyzed to determine if meaningful patterns in the documentation of these errors occurred.

Results: A total of 662 errors were recorded among the 24 candidates. History-taking items were more often misrepresented (n = 378) than physical examination items (n = 284). Patient history—a subcategory that included past medical, surgical, medication, allergy, and family histories—was more likely to be misrepresented than other subcategories under history taking (eg, “history of present illness” and “associated symptoms”) and many physical examination subcategories. Results of the analysis indicated that the misrepresentations were not evenly distributed ($\chi^2_{10} = 323.1; P < .01$): candidates whose SOAP notes contained misrepresentations were statistically more likely to misrepresent in some categories rather than in others.

Conclusion: Physicians and medical students should pay particular attention to patient history, lifestyle history, and neurologic examination to minimize the risk of documentation errors.

doi:10.7556/jaoa.2014.004
In 2007, the National Board of Osteopathic Medical Examiners (NBOME), the agency that administers the COMLEX-USA Level 2-PE, instituted a process to identify candidates who documented history items that were not elicited from the patient or derived from physical examination maneuvers that were performed during the encounter: errors of commission, or “misrepresentation.” Candidates who are found to have repeatedly misrepresented information on their notes are issued failing score reports annotated with “irregular conduct.” Procedures for this policy implementation—including identification of misrepresentation on a SOAP (subjective, objective, assessment, plan) note, the process of subcommittee review and decision making—have been described in the literature.6,9,12,13

The purpose of the present study was to calculate the frequency with which specific items were misrepresented in the SOAP notes and to explore patterns of misrepresentation in the categories of history taking and physical examination. If the areas of the SOAP notes in which candidates are more likely to make errors of commission are identified, this information could prove useful for educational purposes. Medical students, as well as practicing physicians, can be cautioned not only about the danger of making documentation errors of commission, but also specifically where those errors are more likely to occur.

Methods

Examination Format

The COMLEX-USA Level 2-PE is a performance-based examination designed to assess the clinical skills of osteopathic medical students before entry into graduate medical education.6,14 During the examination, candidates rotate through 12 stations wherein they evaluate and treat standardized patients (SPs) who present with various complaints. Encounters are digitally recorded for scoring and quality assurance purposes. Following each 14-minute encounter, candidates are given 9 minutes to

Written documentation of the patient-physician encounter remains an important means to convey health care information to health care professionals, third-party payers, and patients. These groups depend on accurate documentation of the patient-physician encounter, whether written or electronic. If this documentation is incorrect or incomplete, patient care may be adversely affected. Medical errors as a result of breakdowns in the system or in the communication of medical information contribute to a significant percentage of errors that affect patient care, as do cognitive or diagnostic errors.1,2 Furthermore, physician errors in medical diagnosis may occur because of lack of accurate information derived from the patient or from incomplete medical records (eg, laboratory test results) available at the time of decision making. The accurate transfer of medical information—such as history and physical findings—relies heavily on human competency, which may be affected by fatigue, carelessness, and being overworked.1,3,4

In the United States, the skill of documenting patient-physician interaction is assessed by means of 2 clinical skills examinations used for licensure: the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE) and the United States Medical Licensing Examination Step 2 Clinical Skills.5,6 In these examinations—as well as in the educational arena—discrepancies between what is performed in the patient-physician encounter and what is documented on the postencounter patient note have been identified.7,9 These discrepancies are especially concerning given that, on the basis of practice data, inaccuracies in written documentation have contributed to medical errors, which in turn adversely affect patient outcomes.3,5 The errors found in these educational settings have been classified as errors of transcription (incorrect documentation), errors of omission (leaving out items from a written or typed note), or errors of commission (documentation of history not elicited or physical examination maneuvers not performed).7,9,11
complete a SOAP note). Candidates are provided with a SOAP note form and scrap paper in the encounter room. Candidates are scored on 2 domains: the Humanistic domain, which includes patient-physician communication, interpersonal skills, and professionalism, and the Biomedical/Biomechanical domain, which comprises data gathering (history taking and physical examination), SOAP note, and performance of osteopathic manipulative treatment (OMT). The SP encounters for the COMLEX-USA Level 2-PE replicate typical practice patterns of osteopathic physicians. Each SP varies in age, sex, and race/ethnicity. Some SPs present with acute symptoms or describe chronic health issues, and others allow the candidates to discuss issues related to health promotion and disease prevention. Candidates are exposed to content categories, including cardiovascular, respiratory, gastrointestinal, neuro-musculoskeletal, and other (eg, genitourinary, behavioral). The psychometric findings of the COMLEX-USA Level 2-PE scores, detailed elsewhere, are similar to those of other high-stakes clinical skills examinations.

**Misrepresentation of Clinical Findings**

In 2007, the NBOME implemented a process to identify SOAP note misrepresentation. Candidates are informed before the examination that misrepresenting findings in the SOAP note could result in a fail decision and a score annotation of “irregular conduct.” The screening, flagging, and review procedures have been described in a previous study.

Potential SOAP note misrepresentation is identified by raters and by algorithm. Raters are osteopathic physicians who undergo case-specific training and who are familiar with the facts of the case. Therefore, they are able to identify if an error has occurred in documentation (eg, “No allergies” is written on the medical record of a patient with a known penicillin allergy). Raters are expected to flag SOAP notes that have 2 or more discrepancies in either the subjective or objective section. Second, a heuristic algorithm is employed to review records for the given month. Candidates with lower than average data gathering (checklist) scores and higher than average scores on the S, O, or both parts of the SOAP note are most likely to be selected by this process for further review. These candidates’ individual encounter level scores are then reviewed by physician staff. Candidates with disparate scores on individual encounters are noted and only those candidates with encounter-level discrepancies are further screened.

Regardless of the flagging method, a physician staff member reviews a minimum of 3 notes from the flagged candidate and compares the notes from the corresponding video recordings of SP encounters. If 2 of 3 patient notes are found to contain misrepresentation of clinical findings, or if 1 note is found to contain multiple examples of misrepresentation, all 12 SP encounters are reviewed by staff physicians and compared with their corresponding SOAP notes. Then, the candidate records are forwarded to a subcommittee of physicians for adjudication.

The subcommittee reviews the SOAP note and video recording of the SP encounter. When the subcommittee reviews a candidate’s notes, it determines whether the items documented were misrepresented by having been included in the SOAP note. The subcommittee may conclude that an item was documented but that the item’s source was unclear: for instance, a candidate’s question was inaudible, or part of the physical examination was obstructed by a poor camera angle. All 12 SOAP notes are reviewed during the subcommittee meeting (eg, “Case 100: Candidate did not ask for allergy history, or note onset of the pain; heart not auscultated”). The subcommittee then summarizes and takes into account the discrepancies between the video recording and the SOAP note across all 12 encounters. If the committee decides that there was a consistent pattern of misrepresentation, then the candidate is issued a failing score report with an annotation of “Irregular Conduct.”

**Design**

For the present study, we sorted the confirmed misrepresented items for each candidate into 2 categories: history taking or physical examination. Within the history-taking category, items were separated into the following 5 subcategories: (1) history of
present illness (HPI), which included any factors (eg, onset, palliative and provocative factors, quality, severity, temporal factors) directly related to the chief complaint; (2) associated symptoms, or symptoms related to the chief complaint; (3) review of symptoms, or questions about general symptoms unrelated to the chief complaint; (4) patient history (ie, past medical history, past surgical history, medications, family history, and allergies); and (5) lifestyle, including occupation, as well as drugs, alcohol, smoking, and other habits.

In the physical examination section, items were split into 7 subcategories as follows: (1) head, eyes, ears, nose, throat; (2) cardiovascular system, including heart auscultation, carotid examination, and pulses; (3) pulmonary system; (4) gastrointestinal system; (5) genitourinary system; (6) neurologic examination, including reflexes, cranial nerves, and mental status; and (7) musculoskeletal system. Three physician-examiners (J.M.S., L.A.G., and another physician) participated in the categorization of these items and classified them by consensus. We totaled the number of times that a candidate misrepresented in each of the 12 categories across all 12 notes.

A waiver for Institutional Review Board approval for this project was obtained through the Center for the Advancement of Healthcare Education and Delivery.

**Data Analysis**

We conducted $\chi^2$ tests of independence, with statistical significance set at $P<.01$, to test the hypothesis that misrepresentation occurred equally often in all categories.

**Results**

Among 12,510 candidates from 3 test cycles (2007-2008, 2008-2009, 2009-2010), 24 received a failing score on the COMLEX-USA Level 2-PE because of misrepresentation of clinical findings. From this candidate population, we classified 662 observations of misrepresentation into the 12 aforementioned categories. Per candidate note, the number of items misrepresented ranged from 0 to 19 for each subcategory. Each category could have multiple iterations of misrepresentation (eg, if a candidate misrepresented both alcohol and smoking history, this candidate would be marked as having 2 errors in the lifestyle subcategory of history taking).

The number of items misrepresented on candidates’ notes ranged from 12 to 60 items. History-taking items were more commonly misrepresented ($n=378$) compared with physical examination items ($n=284$). As seen in the Table, the number of commissive errors ranged from 2 (genitourinary system) to 151 (patient history). Misrepresentations were not evenly distributed ($\chi^2=323.1, P<.01$); candidates whose notes contained misrepresentations were statistically more likely to misrepresent in some categories rather than in others.

Because individual cell contributions to the overall $\chi^2$ value are themselves distributed $\chi^2$ on 1 degree of freedom, we were able to evaluate each category to investigate if the misrepresentation rate within it was significantly higher than expected. (Misrepresentation rates that were significantly lower than expected were not evaluated, except to note that misrepresentations in the pulmonary category were relatively rare.) Three categories—patient history ($N=151; \chi^2=138.02, P<.01$), neurologic examination ($N=117; \chi^2=54.15, P<.01$), and lifestyle ($N=99; \chi^2=25.35, P<.01$) — contained a disproportionately high number of misrepresentations.

The Figure illustrates the extent to which the frequency of errors in each category was greater or less than would be expected given the average number of errors per category. The x-axis is in standardized residual units, which are calculated as the square root of the individual category contribution to the overall $\chi^2$. The Figure also illustrates the overrepresentation of patient history, neurologic examination, and lifestyle, the front-runners in misrepresentation frequency; review of systems is revealed as a distant fourth place.

We conducted another $\chi^2$ analysis to establish whether candidates differed in the number of categories in which misrepresentations occurred. Genitourinary misrepresentations were included in the count for this analysis. The number of categories in which misrepresentations occurred ranged from 3 to 11, with a mean (standard deviation) of 6.625 (1.97), indicating that candidates who
Nevertheless, human error may contribute to an inaccurate recording of findings or events that did not transpire (e.g., 1 patient confused with another). The present study investigated if commissive errors occur with greater frequency in some parts of the patient history and physical examination than in others. Physicians and trainees may then be especially careful to document these parts of the encounter accurately.

During the COMLEX-USA Level 2-PE, commissive errors were found to occur more frequently in the history-taking category compared with the physical examination category. This finding may be the result of a variety of factors, including behavioral ones. In some instances, candidates may confuse the history of the patient just seen with that of 1 cared for earlier in the day. In other instances, physical examination may be easier to accurately remember because the candidate actively participates in the process. Also, it may be that some candidates seem to follow a formulaic history-taking agenda, from which they may be diverted by something the patient says (e.g., an interesting finding, a shared experience). When candidates are diverted from that blueprint, they may forget that they did not ask something and document their usual history and examination categories.

In the subjective portion of the SOAP note, items from the patient history subcategory were most often misrepresented on candidate notes, followed by items in the lifestyle subcategory. These items were more likely to be misrepresented than others from the history-taking category (history of present illness, associated symptoms), as well as many systems from the physical-examination category (musculoskeletal, gastrointestinal, pulmonary, and genitourinary). Because they are almost always elicited, items in the patient history subcategory are more easily confused from patient to patient. Further, the formulaic way in which the items are elicited may cause a candidate to overlook a particular item as he or she comes across an interesting part of the patient’s history. Other subcategorical items—such as history of present illness or associated

### Table.
Categories of 662 Commisive Errors Made by 24 Candidates During the COMLEX-USA Level 2-PE

<table>
<thead>
<tr>
<th>Category</th>
<th>Errors, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>History Taking</strong></td>
<td></td>
</tr>
<tr>
<td>History of present illness</td>
<td>31 (4.7)</td>
</tr>
<tr>
<td>Associated symptoms</td>
<td>29 (4.4)</td>
</tr>
<tr>
<td>Review of systems</td>
<td>68 (10.3)</td>
</tr>
<tr>
<td>Patient history*</td>
<td>151 (22.8)</td>
</tr>
<tr>
<td>Lifestyle*</td>
<td>99 (15.0)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>378 (57.1)</td>
</tr>
<tr>
<td><strong>Physical Examination</strong></td>
<td></td>
</tr>
<tr>
<td>Head, eyes, ears, nose, throat</td>
<td>44 (6.6)</td>
</tr>
<tr>
<td>Cardiovascular system</td>
<td>48 (7.3)</td>
</tr>
<tr>
<td>Pulmonary system</td>
<td>12 (1.8)</td>
</tr>
<tr>
<td>Gastrointestinal system</td>
<td>27 (4.1)</td>
</tr>
<tr>
<td>Genitourinary system</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>Neurologic examination</td>
<td>117 (17.7)</td>
</tr>
<tr>
<td>Musculoskeletal system</td>
<td>34 (5.1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>284 (42.9)</td>
</tr>
</tbody>
</table>

* Includes past medical history, social history, medications, allergies, and family history.

* Includes social history.

**Abbreviation:** COMLEX-USA Level 2-PE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation.
symptoms—vary from patient to patient and are more likely directly related to the reason for the visit, perhaps making them more immediate, more memorable, and therefore more accurately documented as a result (eg, the presence of fever in a child with an upper respiratory infection vs the child’s family history). We were a bit surprised that review of systems was in a distant fourth place; because review of systems encompasses so many areas, we assumed it would be more often forgotten during history taking or documentation.

Collecting the patient history and lifestyle information is essential to uncovering more about the “whole patient” and is therefore an integral part of the osteopathic physician’s encounter with a patient. The first osteopathic tenet tells us that the patient is a unit of body, mind, and spirit.\textsuperscript{15} The greater number of errors in patient history and lifestyle subcategories could reflect that students who forget to elicit these items during the encounter later recognize these items or areas as important when writing the note and document them there (either thinking that they did ask or knowing that they should have). Similarly, the neurologic examination is important to the osteopathic examination and the body’s structure-function relationship, and the same principle—that students may recognize their importance, but too late—could apply to why these errors are more prevalent here.

In the objective section of the SOAP note, only the commissive errors in the neurologic examination were identified as disproportionately high. The reason for this finding is unclear, but we suspect that because many of these items are documented in groups (eg, cranial nerves, reflexes), students may have performed only part of a physical examination but documented that a complete examination was performed.

---

**Figure.**
The frequency of commissive errors in 11 subcategories for 24 candidates on the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation. The candidates were drawn from a total of 12,510 candidates in 3 testing cycles (2007-2008, 2008-2009, and 2009-2010). The frequency was greater or less than would be expected given the mean number of errors per category. The x-axis is in standardized residual units, which is calculated as the square root of the individual category contribution to the overall $\chi^2$ analysis. The subcategory “lifestyle” included social history. The subcategory “patient history” included such items as past medical history, social history, medications, allergies, and family history.
Moreover, our results suggest that misrepresentation seemed to occur in patterns in these candidates (ones who demonstrate misrepresentation repeatedly), but this finding would need to be confirmed by a larger sample of students.

Limitations
The present study was limited by several factors, including a small study population, a limited error type, the natural constraints of the examination, and oversampling.

First, there were a small number of candidates because we reviewed only those who had a consistent pattern of these types of errors on their notes (enough to lead to a failing score being issued for the examination). Therefore, the study may not generalize to candidates with fewer errors. Second, we confined the present study to 1 of 3 possible errors (commissive errors), whereas it could be argued that the reason for a mistake in documentation does not matter if patient care was compromised. The NBOME issues these types of failing score reports only to those candidates who have a notable pattern of these types of errors on their notes over an entire test day. The review process only flags and identifies these errors. A study looking at all errors on notes regardless of outcome would be interesting and may provide additional information.

Furthermore, some physical examination maneuvers are prohibited in this examination (ie, rectal examination, female breast and pelvic examination, male genital examination, and corneal reflex testing, as well as OMT techniques such as articulatory and high-velocity, low-amplitude), which may affect these results. In addition, the performance of OMT and the osteopathic structural examinations were classified with the appropriate subcategory (eg, sinus drainage was associated with head, eyes, ears, nose, and throat; facilitated positional release to the thoracic spine was associated with the musculoskeletal system). The exclusion of the prohibited actions could potentially increase the rate of misrepresentation in certain subcategories more than others. The low error rate in the genitourinary category is likely most affected by these prohibited techniques—there is little opportunity for these types of examinations to be performed.

Finally, we may be oversampling from the subjective section in the initial screening because these are more easily identified as errors by raters. As a rater, identification of misrepresentation in the objective section of the note is challenging because there tends to be less variability in this section in SP examinations.

Issues With Electronic Medical Records
Commissive written errors—the focus of the present study—are only 1 potential source of medical errors on a clinical skills examination. At present, electronic health records and electronic prescribing applications are regarded as tools that may reduce errors throughout the health care system.4,16 Some studies,17-20 however, show disappointing results with respect to the association between electronic medical record use and overall patient care outcomes. Nonetheless, an accurate medication history—for example, including adverse reactions to medications and the steps that follow (eg, prescribing, dispensing, administrating)—all depend on the accuracy of the initial record. The use of electronic medical records may, unfortunately, open up the possibility of increased errors in the record through the inattentive use of check boxes and the copy-and-paste function.20,21 Clinicians need to be vigilant about how they document information from a patient encounter to make sure that any record, whether written or electronic, is an accurate reflection of the patient-physician encounter.

With the advent of high-stakes clinical skills testing for licensure in 2004, the foundational skills of medical history-taking, physical examination, and patient-physician communication have been emphasized more and more in the undergraduate medical education setting. As has been pointed out by colleagues in previous studies,7-9 documentation of the patient-physician encounter should be reinforced during the clinical years in the interest of patient safety. The present study may show educators areas where students’ attention can be focused or, at the very least, areas in which a student’s caution should be directed.
Conclusion

Patient history, lifestyle history, and neurologic examination are areas in the SOAP note where commissive types of errors in documentation may occur more frequently during a clinical skills examination. Although all aspects of a patient-physician encounter are important, educators should direct osteopathic medical students to pay particular attention to these areas. Care should be taken throughout medical education and beyond to minimize the risk of any documentation errors and to improve patient care.

References


