Evaluating Physician Performance at Individualizing Care: A Pilot Study Tracking Contextual Errors in Medical Decision Making

Saul J. Weiner, MD, Alan Schwartz, PhD, Rachel Yudkowsky, MD, MHPE, Gordon D. Schiff, MD, Frances M. Weaver, PhD, Julie Goldberg, PhD, Kevin B. Weiss, MD

BACKGROUND

Clinical decision making requires 2 distinct cognitive skills: the ability to classify patients’ conditions into diagnostic and management categories that permit the application of research evidence and the ability to individualize or—more specifically—to contextualize care for patients whose circumstances and needs require variation from the standard approach to care. For instance, intensifying the insulin regimen in a diabetic patient with worsening glucose control reflects the former; adapting an alternative approach when the problem is linked to failing eyesight, the loss of a caretaker who assisted with medication administration, or deteriorating cognitive status reflects the latter.

A clinical decision that is sufficiently misguided may result in a medical error, defined as “the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim.” In the example above, errors might include reducing insulin erroneously when the patient is already on too low a dose or, alternatively, prescribing a higher dose when the problem is that he or she can no longer follow directions. The former has been classified as a biomedical error and the latter a contextual error, and data were collected on information elicitation and error making. Results. The case validation process was successful in that, in the final iteration, the physicians who received the contextual variant of cases proposed an alternate plan of care to those who received the baseline variant 100% of the time. The subsequent piloting of these validated cases unmasked previously unmeasured differences in physician performance at contextualizing care. The findings, which reflect the performance characteristics of the study population, are presented. Conclusions. This pilot study demonstrates a methodology for measuring physician performance at contextualizing care and illustrates the contribution of such information to an overall assessment of physician practice. Key words: medical error; patient-centered care; communication; quality of care. (Med Decis Making 2007;27:726–734)

Received 28 February 2007 from the Department of Medicine (SJW), Department of Pediatrics (SJW, AS), and Department of Medical Education (AS, RY, JG), University of Illinois at Chicago; the Veterans Affairs Center for Management of Complex Chronic Care, Chicago, Illinois (SJW); John H. Stroger Jr. Hospital of Cook County and Rush University, Chicago, Illinois (GDS); Northwestern University Feinberg School of Medicine, Chicago, Illinois (FMW, KBW); and the Center for Management of Complex Chronic Care, Hines VA Hospital, Hines, Illinois (FMW, KBW). Source of support: Robert Wood Johnson Generalist Physician Faculty Scholars Program and the Veterans Health Administration. Revision accepted for publication 4 April 2007.

Address correspondence to Saul J. Weiner, MD, Department of Medicine, University of Illinois at Chicago, 840 South Wood Street, M/C 718, Chicago, IL 60612; e-mail: sweiner@uic.edu.

DOI: 10.1177/0272989X07306113
error. Biomedical errors occur because of inattention to processes occurring under the skin (such as the effect of insulin on glucose metabolism), and contextual errors are due to inattention to processes expressed outside the skin—that is, processes that form the context of a patient’s illness (such as the effect of poor vision or cognitive function on insulin self-administration), even when they originate from a biomedical condition.

Avoiding such errors involves appropriately adapting research evidence to patient context so that a clinician can recognize and discuss the full implications of various treatment strategies. The patient is then in a position to express his or her preferences based on all relevant information. Of course, patient context does not always require a deviation from standard approaches to care. Clinical decisions often require only biomedical information. However, a consideration and exploration of context must be an essential component of each clinical encounter given the potential clinical significance of such information. Contextualizing decision making requires the motivation and cognitive skills to seek out factors unique to each patient’s life situation relevant to planning his or her care.

Assessing physician performance at contextualizing care presents distinct challenges. Although there are evidence-based guidelines for determining, for instance, whether a clinician has correctly adhered to recommendations for anticoagulation in a patient with lone atrial fibrillation, how does one assess whether the physician recognizes when those guidelines may be inappropriate in the setting of contextual factors, such as transportation difficulties that might complicate safe administration and monitoring of the medication? To appreciate the challenges of such an assessment, consider the limitations of the commonly used approach of direct observation by a teaching clinician charged with evaluating the performance of a resident. First, there is the issue of ascertaining the validity of the teacher’s determination about what constitutes appropriate care. The problem is that there can never be research evidence to demonstrate the impact of an individual’s unique context on outcome for any clinical condition because every individual’s context is slightly different. When a resident and attending disagree about whether to recommend for or against anticoagulation based on a particular patient’s life situation, on what grounds is the attending’s conclusion deemed more valid than the resident’s? Clearly, there are situations when following recommendations for anticoagulation would, in fact, pose serious risks that outweigh benefits, under specific circumstances. In assessing a physician’s performance, however, how can we be confident in the superior judgment of the experienced attending?

The second challenge involves risk adjustment, an issue that arises when attempting to compare clinicians’ performance. The problem is that no 2 clinical dilemmas are alike when the relevant variables are contextual factors that make them distinct. The risks and implications of falling will never be quite the same for 2 individuals taking warfarin because of numerous circumstantial variables (e.g., whether there are loose carpets in the home, slippery steps, whether the patient lives alone or with someone else, has good vision, etc.), even when the biomedical aspects of the cases are the same. How, then, can one compare the recommendations of 2 clinicians when the information on which those recommendations are based differs? Again, the situation contrasts with strictly biomedically based decision making, where the focus is on discrete evidence-based criteria (e.g., how long a patient has been in atrial fibrillation) that allow for comparisons across multiple providers.

The third problem involves clinical authenticity, which arises whenever one assesses clinical performance using written scripted cases where the necessary information is fully provided rather than elicited. If one were to design a case scenario, for instance, in which clinicians were presented with all the information about a patient who might require anticoagulation, including his or her history of previous falls, bleeds, cognitive deterioration, and poor supervision, it would become apparent to many that the intended “correct answer” is to recommend against anticoagulation. However, under “real-life” conditions, those same providers might never think to elicit the critical contextual information.

The purpose of this study is to develop and test a methodology for measuring physicians’ performance at contextualizing care and compare it to their performance at planning biomedically appropriate care. We describe how our findings have led to a larger study to explore potential moderators of contextual and biomedically based errors in the medical decision-making process.

**METHODS**

We set out to construct and test clinical cases that would overcome the aforementioned challenges to assessing the performance of a group of clinicians at
identifying and making use of contextual information. First, each would incorporate contextual factors so overwhelmingly significant to planning patients’ care that the introduction or removal of those factors would predictably alter what seasoned clinicians would consider appropriate care. Such cases would serve as valid instruments for assessing clinical performance.

The second objective was to eliminate the variability in context across patients with the same biomedical condition. This could be achieved through the use of standardized patients (SPs) as opposed to other methods of practice evaluation such as chart audits or even direct observations of practice behavior. Standardized patients are inherently risk adjusted because each clinician gets the same scenario. Standardization is essential for exploring variability in individualized decision making because it enables comparison of how different providers respond to the same life narrative at a particular point in time.

The third objective was to present the cases to clinicians so that the critical contextual information would never be offered voluntarily but would have to be elicited. Such an approach would overcome the problem of clinical authenticity. Clinicians would have to demonstrate that they pursue contextual factors during encounters, rather than simply react to information given to them.

In our analysis of the data generated from piloting these cases, we sought to isolate trainees’ skills at biomedical versus contextual reasoning to enable comparisons of their performance in these 2 areas as a way of “diagnosing” specific causes of performance failure in the biomedical and contextual aspects of clinical decision making.

Case Development

To achieve these aims, we developed 4 scripted cases, each of which has 4 variations, based on actual patient encounters in the ambulatory setting. Each begins with a straightforward clinical problem, typical of what a primary care physician might see in the office setting. The script of each case, however, has 4 variations: baseline, baseline + biomedical qualifier, baseline + contextual qualifier, and baseline + biomedical + contextual qualifier. We have coined the term biomedical qualifier to describe the physiological red flags that, if elicited, point the physician in a new direction and, if not elicited, lead to a biomedical error. We use the term contextual qualifier to describe the contextual red flags that, if elicited, should point the clinician in a new direction (i.e., one requiring a different approach to management to avert a contextual error). Three of the 4 cases were selected for piloting in this study. Their essential components are summarized in Table 1.

In all versions of a case, the SP is instructed to reveal a narrative that suggests a typical solution to a clinical problem. In the baseline version, questioning yields no additional relevant information. This version of the case has no surprises; it is entirely routine. In the biomedical version, however, although the presentation is the same as baseline, if appropriate questions are asked, an atypical biomedical narrative will unfold, suggesting an alternative diagnosis. The new alternate diagnosis requires a different approach to management. In the contextual version, the presentation is, again, the same as in the baseline; however, this time, appropriate questions will elicit information about the patient’s life situation that would change how the case should be managed. Finally, in the biomedical + contextual version, appropriate questioning will elicit both qualifiers, leading to both the atypical biomedical diagnosis and the confounding patient circumstance. An appropriate plan of treatment should take both of these factors into account.

A critical element of the case design is that both the contextual and biomedical qualifier information be essential to patient management rather than simply useful or beneficial. A helpful analogy is a classification scheme used to grade diagnostic errors: “class 1 error,” defined as “a major unexpected finding . . . that would lead to a change in therapy and improved survival” is differentiated from “class 2 error,” where inattention to major unexpected findings would not affect management. Our intent was to design the cases so that all errors, biomedical and contextual, would misdirect management, rendering them class 1 errors. The clinical significance of the biomedical qualifiers was based on research evidence and the standard of care. The clinical significance of the contextual qualifiers was based preliminarily on our own assessments and, subsequently, on a validation process performed by physician peers who reviewed the cases.

The Validation Process

As noted above, once the cases were drafted, they underwent a validation process, during which time they were refined and established as instruments for assessing clinician performance. The first step
involved informally presenting each narrative in written form to internal medicine colleagues, both with and without the biomedical and contextual qualifier information, and asking them to recommend a plan of care. This was an iterative process in which we revised the text until the biomedical and contextual qualifier information consistently prompted different approaches to management. Next, we presented 16 different practicing board-certified general internists with a text version of 1 variant of each of the cases, posted on a Web site, with each participant provided an e-mail link to his or her randomly assigned set. The combination of variants counterbalanced in a fractional factorial design across subjects. The physician subjects were instructed to “strive for optimal care, but avoid recommendations that are not necessary for optimal care.”

The cases were considered valid instruments for assessing the quality of physician performance when 4 out of 4 physicians who had never seen previous iterations of a case variant independently reached agreement about how biomedical or contextual qualifier information would alter their plan of care. An example of how a case variant was presented, and then coded, is illustrated in Box 1. The exercise was approved as exempt from review by the Institutional Review Board at John H. Stroger Jr. Hospital of Cook County.

Box 1
An Illustration of the Case Validation Process

The following example illustrates how case Amc (i.e., the version with both biomedical and contextual qualifiers) was presented on the Web site and a physician subject’s response. For illustrative purposes below, we have labeled and separated the baseline, contextual, and biomedical qualifier information.

**Baseline:** Mr. James is a 42-year-old man who comes to see his physician about persistent respiratory difficulty that he believes is related to his asthma; he has never smoked. He reports that he was diagnosed with asthma as a child but rarely had problems until about 5 years ago, when he began to experience more frequent occurrences of shortness of breath and wheezing. These episodes, which occur every few days, are relieved with the use of an albuterol inhaler. He and his wife now have 3 young children, ages 5, 3, and 2, who often have colds that get “passed around” the family. Recently, a physician ordered pulmonary function tests that demonstrated reversible airway obstruction, confirming the diagnosis of...
reactive airway disease. Mr. James reports that he was
told he would “need one of the steroid medications”
and was prescribed Pulmicort Turbuhaler, 1 puff bid.
He reports that it has helped but that he is still having
problems with wheezing periodically.

**Contextual qualifier:** However, if asked about how
or when he takes his medication, Mr. James will state
that he “doesn’t like taking his medication regularly.”
When this is pursued, he will reveal that he has
heard that steroids are a hormone that can hurt you
and that he thinks it may be causing him to have trou-
ble with erections. He will acknowledge that he has
been using his Flovent only when his symptoms are
bad, not daily, because of these concerns.

**Biomedical qualifier:** When asked about the his-
tory of his symptoms or exacerbating conditions, Mr.
James reveals that his symptoms of wheezing are
worse since he gained weight, that they occur more
frequently when he is supine after eating large fatty
meals, that he periodically experiences some burning
in his chest that he treats with Tums, and that his
voice is periodically hoarse. Further questioning does
not lead to any more clinically revealing information.

**Physical exam:** Other than an occasional scattered
wheeze, the physical exam is unremarkable.

One physician subject’s response (with our com-
ments added in italics):

Clearly Mr. James has URI triggered asthma and
would benefit from an inhaled steroid. Also he has
GERD symptoms that could be exacerbating his
bronchospasm (biomedical qualifier acknowl-
edged by subject). Since Mr. J is not taking his
medication daily because he has confused it with
anabolic steroids (contextual qualifier acknowl-
edged) I would explain to him the difference and
reassure him of its safety. I would also talk with
him about his erectile dysfunction (contextual
qualifier deemed necessary to plan of care). For
his GERD I would put him on a PPI (biomedical
qualifier deemed necessary to plan of care).

The physician’s response was coded using the fol-
lowing checklist developed for the case, with the
coder blind to the case variant he was coding (our
comments in italics added, below):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>
| The physician recommends stepped-up asthma therapy with addition of a long-acting beta-
  agonist, a leukotriene inhibitor, or a higher dosage of inhaled glucocorticoid. (This would be
  the correct management for baseline variant.)                            |     |    |

Note that the “correct management” of biomedical
information is based on research evidence and the
standard of care. The correct management of contex-
tual information, specific to each case, was proposed
prior to this validation study and then validated
when all physicians independently deemed the
information necessary to care.

In the Amc case, 4 out of 4 physicians’ responses
were coded identically, rendering the case valid as
an instrument for assessing the quality of physician
practice. The proposed correct management of the
contextual information became, then, the valid cor-
correct management based on consensus. Any physician
who subsequently fails to address the patient’s misin-
formation about the medication will be considered to
have made a contextual error.

**Piloting of the Validated Cases**

**Using Standardized Patients**

Three actors were trained as standardized patients
to present 1 case and its variants to 54 internal medi-
cine residents during an unrelated mandatory educa-
tional activity organized by the University of Illinois
at Chicago Clinical Performance Center (CPC) from 7
January to 4 March 2005. The SPs were instructed to
answer all questions in a forthright manner during
their encounters. The CPC is a specialized facility
for SP training and testing and uses a standard proto-
col for verifying the performance and consistency
of its actors (http://www.uic.edu/com/cpc/revised
_03). Approximately 6 hours of training is provided
for each case, during which time actors are video-
taped and scored by trainers until they achieve
100% accuracy in 1) presenting the baseline clinical...
information (eliminating the need for risk adjustment) and 2) revealing the additional biomedical and/or contextual qualifier information only when asked (to achieve clinical authenticity). Twenty percent of the actual encounters were audited for any decay in performance.

Each resident was instructed to visit the CPC for an afternoon to interview 6 SPs as part of their training, 1 of whom presented a case variant for this study. Twenty-five residents received a version of case A, 12 a version of case B, and 17 a version of case C (variation due to the availability of the actors). Residents were assigned to sessions based on their call schedule. Exactly half were PGY 2 and half were PGY 3 residents, and 55% were men. The CPC has approval from the institutional review board of the University of Illinois for tests of the performance characteristics of cases for educational and research purposes.

Two checklists were developed for each patient case, one to be completed by the SP after the encounter and the other by a coder reviewing the physician’s notes. The SP documented elicitation (or nonelicitation) of the biomedical and contextual qualifier information for each case. For example, for the case of Mr. James, the SP documented whether he revealed, based on a closed-ended question by the physician, 1 of 4 biomedical “red flags”: heartburn, hoarseness, reduced heartburn with Tums, or worsening symptoms when supine after large fatty meals. In addition, the SP indicated if he revealed one of the “red flags” in response to an open-ended question (to reward both open-ended as well as closed-ended questions). A parallel checklist was used for contextual “red flags.” Figure 1 illustrates the checklist for the James case.

We designed a second checklist to differentiate physicians who incorporated into their management plan (after eliciting) the clinically relevant biomedical and contextual information embedded in each case from those who did not. It is based on the checklist developed in the case validation exercise described above and was coded by a physician investigator blind to knowledge of the case variant. For the James case, the coder documented whether the physician recommended stepped-up asthma therapy (correct for the baseline), a proton pump inhibitor (PPI), an H2 blocker, or further evaluation for gastroesophageal reflux disease (GERD; correct for the biomedical variant) or educated the patient about the difference between glucocorticoids and other steroids in the setting of the patient’s concerns about erectile dysfunction (correct for the contextual variant). An example of how the James biomedical + contextual version would be coded if managed correctly is illustrated in the table in Box 1.

The primary outcome variables were successful elicitation of information available (i.e., whether biomedical and contextual information were elicited when present in the variant) and successful incorporation of elicited information available (i.e., whether the resident chose the correct plan of care given the information in the variant).

All information was entered directly into the CPC’s Web-based software for SP data tracking and

---

**Figure 1** Coding checklist for elicitation of information for all 4 variants of the “James case.” Note that for the baseline version of the case, the standardized patient (SP) will always answer no to all. For the biomedical variant, he may answer yes to 1–5; for the contextual variant, he may answer yes to 5–10; and for the biocontextual variant, he may answer yes to all.
exported into SPSS (version 13) or SAS (version 8.2) for analysis. Mixed logistic regression models were used to compare 1) the rate of eliciting biomedical information with the rate of eliciting contextual information in case variants where either or both were present and 2) the rate of incorporation of elicited biomedical information with the rate of incorporation of elicited contextual information in case variants where either or both were present.8

Models were fitted using a generalized estimating equations approach with a logit link function. In the first model, the outcome variable was successful elicitation of information available, as predicted by a single dichotomous variable coding type of information available in the variant (biomedical or contextual). When a variant included both biomedical and contextual information, each was treated as a separate unit in the analysis, and subjects were clustered within units; the working correlation matrix was modeled with compound symmetry. This model allows us to compare the probability of eliciting available biomedical information with the probability of eliciting available contextual information, while accounting for intrasubject correlation among subjects in variants where both kinds of information were available for elicitation. In the second model, the outcome variable was successful incorporation of elicited information available; all other features of the model followed the approach in the first model.

In addition, chi-square tests were used to compare rates of elicitation and incorporation of elicited information among cases (i.e., James, Collas, and Davis), collapsing across variants, to assess case specificity.

RESULTS

Test of Cases Using Standardized Patients

The results of the experiment using standardized patients are illustrated in Table 2. Twenty out of 54 residents (37%) failed to elicit essential qualifier information, 9 from a failure to elicit biomedical information from a biomedical variant, 4 from a failure to elicit contextual information from a contextual variant, and 7 from nonelicitation of either biomedical and/or contextual information from a biomedical + contextual variant. Logistic mixed model regression predicting likelihood of elicitation from the presence of information for these 3 case variants found that residents were significantly more likely to elicit contextual information when it was present than biomedical information (odds ratio [OR] = 4.56, 95% confidence interval [CI] = [1.44, 14.47], $\chi^2(1) = 6.63, P = 0.01$).

Although contextual information was more often elicited, it was less often incorporated: logistic mixed model regression predicting the likelihood of incorporating elicited information from the presence of information—again from all 3 case variants containing qualifier information—found that residents were significantly less likely to incorporate contextual information when it had been elicited than biomedical information (OR = 0.16, 95% CI = [0.027, 0.903], $\chi^2(1) = 4.30, P = 0.038$). For example, although there were only 4 out of 15 failures to elicit contextual information in the contextual variant, 9 of these resident physicians had incorrect care plans (i.e. made contextual errors), indicating the occurrence of 5 failures to incorporate successfully elicited contextual information. In the biomedical variant, no errors occurred above and beyond those following the 9 out of 14 failures to elicit essential biomedical information. That is, when biomedical information was elicited, it was always incorporated into the plan.

There was evidence of case-specific differences in performance in incorporation but not elicitation. Residents were equally likely to elicit the available information in each case ($\chi^2(2) = .912, P = 0.63$) but were not equally likely to incorporate elicited information ($\chi^2(2) = 17.7, P < 0.001$). Residents were less likely to incorporate elicited contextual information in the Davis case than in either the James or Collas case.

DISCUSSION

This study takes a first step toward explicitly characterizing and quantifying individualized care. It does so by introducing a methodology for exposing “contextual error,” a form of medical error that results from a failure to take into account critical contextual information when planning a patient’s care. We operationalized the concept first, by demonstrating that seasoned clinicians are able to reach the same high consensus regarding what constitutes essential contextual information as they can regarding what constitutes essential biomedical information relevant to medical decision making. To achieve this consensus, we selected situations where the significance of a patient’s life situation to his or her care plan appeared self-evident. There are, of course, many more instances where the implications of
context are far less clear. The same can be said for biomedical information. Not all inattention to biomedical data constitutes error. In both cases, the threshold is “the use of a wrong plan to achieve an aim.”

We then tested our methodology on a group of resident trainees using the validated cases and found that—in our particular study population—biomedical and contextual errors were about equally common, but not for the same reasons. The physicians studied were more likely to identify essential contextual information than biomedical information but less likely to make use of the former when planning patient care. No one, for instance, who identified signs and symptoms of hypothyroidism in Ms. Collas neglected to treat the condition. However, several physicians, after learning that she had sole caretaker responsibility for a dying child, neglected to address the possibility of her postponing elective surgery. These findings challenge the view that physicians fail to address psychosocial issues because they have not heard them. SPs are uniquely valuable for comparing physician performance because they are intrinsically risk adjusted: they present the same scenario over and over again. For this study, the SPs were instructed always to be forthright in responding to all questions (i.e., never to be coy or indirect). Hence, we suspect that our methodology would more likely underrepresent the true failure rate of physicians to elicit important information relevant to patient care in actual patient encounters.

To address these limitations, we are enrolling a large cohort of fully trained primary care clinicians in office practices and employing undercover “unannounced” standardized patients with hidden audiotape recorders. Data will also be collected on clinician characteristics (e.g., training, experience, sociodemographics), encounter conditions (e.g., length of visits, presence or absence of an electronic medical record), communication dynamics (through interaction analysis), and patient-specific variables (e.g., each case will have an African American and a Caucasian actor) to canvas broadly for potential moderators of physician error making. Our goal is to understand the causes of variability in clinician performance. The acquisition of such basic knowledge may be essential to improving the medical decision-making process in situations where appropriate care critically depends on taking into account a patient’s unique circumstances and needs.

In summary, improving the quality of physician decision making depends on valid metrics for measuring performance. To date, assessments of decision making have focused on the processing of biomedical information, primarily through gauging adherence to evidence-based practice. A recent

### Table 2

<table>
<thead>
<tr>
<th>Variant</th>
<th>Sample Size</th>
<th>Failed to Elicit Information (%, [95% CI])</th>
<th>Incorrect Plan of Care (%, [95% CI])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1</td>
<td>12</td>
<td>(nothing to elicit)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Biomedical</td>
<td>14</td>
<td>9 (64, [38, 84])</td>
<td>9 (64 [38, 84])</td>
</tr>
<tr>
<td>Contextual</td>
<td>15</td>
<td>4 (27 [11, 52])</td>
<td>9 (60 [35, 80])</td>
</tr>
<tr>
<td>Biomedical and contextual</td>
<td>13</td>
<td>7 (54 [29, 77])</td>
<td>8 (62 [35, 82])</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>20 (37 [25, 50])</td>
<td>26 (48 [35, 61])</td>
</tr>
</tbody>
</table>

*In the variant with both biomedical and contextual information, 5 respondents failed to elicit biomedical information, 1 respondent failed to elicit contextual information, and 1 respondent failed to elicit either biomedical or contextual information. Two respondents had a plan of care that was biomedically incorrect, 2 had a plan of care that was contextually incorrect, and 4 respondents had a plan of care that was both biomedically and contextually incorrect.*

DECISION MAKING AND JUDGMENT

733

Downloaded from mdm.sagepub.com at Society for Medical Decision Making Member Access on March 14, 2011
analysis of the literature on physician evaluation concluded that “methods of assessing physicians’ practice performance are in their infancy, and face unresolved logistic and psychometric challenges.” Developing an empirical method for identifying contextual error makes it possible to begin to explore the causes of these errors that have previously eluded rigorous analysis.

REFERENCES