

Relating Faults in Diagnostic Reasoning With Diagnostic Errors and Patient Harm

Laura Zwaan, MSc, Abel Thijs, MD, PhD, Cordula Wagner, PhD, Gerrit van der Wal, MD, PhD, and Daniëlle R.M. Timmermans, PhD

Abstract

Purpose

The relationship between faults in diagnostic reasoning, diagnostic errors, and patient harm has hardly been studied. This study examined suboptimal cognitive acts (SCAs; i.e., faults in diagnostic reasoning), related them to the occurrence of diagnostic errors and patient harm, and studied the causes.

Method

Four expert internists reviewed patient records of 247 dyspnea patients, using a specially developed questionnaire to detect SCAs. The patients were treated by 72 physicians between May 2007 and February 2008 in five Dutch hospitals.

The findings of the record review were discussed with the treating physicians, and the causes of SCAs were classified using Reason's taxonomy of unsafe acts. Statistical analyses were performed with descriptive statistics and independent *t* tests to compare groups. Furthermore, a reliability study was conducted to assess the interrater reliability.

Results

SCAs occurred in 163 of 247 cases reviewed (66%). In 34 (13.8%) of all cases, a diagnostic error occurred, and in 28 (11.3%) cases, the patient was harmed. Cases with diagnostic errors or patient harm had more SCAs. However,

in 10 (4.0%) of the cases, diagnostic errors or patient harm occurred, though there were no SCAs. The causes of SCAs were mostly mistakes (i.e., the planned action was incorrect).

Conclusions

In cases with more SCAs, diagnostic errors and patient harm occurred more often, suggesting that the number of SCAs per case was predictive of the occurrence of these events. The most common causes were mistakes, meaning that physicians did not realize their actions were incorrect.

Editor's Note: A commentary on this article appears on page 135.

A substantial number of medical errors are classified as diagnostic errors,¹⁻⁶ defined as diagnoses that were unintentionally delayed, wrong, or missed.⁷ Diagnostic errors are often judged to be preventable and have severe consequences.¹⁻⁵ Reported incidence rates of diagnostic errors vary greatly among different studies,^{5,8,9} in part because of variability among research methods. Some studies focus on measuring the outcome of the diagnostic process, such as diagnostic errors or patient harm, but they do not examine the reasoning process behind the outcomes.^{3,5,8,10} Other studies focus on

the reasoning process, detecting different reasoning errors that physicians are subject to during the diagnostic process.^{11,12} However, it is not clear to what extent these reasoning errors eventually led to a diagnostic error or patient harm. Studies that investigate the relationship between the diagnostic reasoning process and diagnostic errors and patient harm are currently lacking.⁹

Establishing such a connection between the diagnostic reasoning process and diagnostic errors also enables the identification of the causes of diagnostic errors, which is important for improving the diagnostic process. Previous studies have used retrospective record reviews to identify causes of diagnostic errors, such as a lack of consultant involvement,¹³ lack of knowledge,⁵ and inadequate transfer of information.⁵ The drawback of these studies is that although the researchers identified why the errors occurred, they did not extensively examine the reasoning processes that led to the diagnostic errors.

To address this, we examined the relationship between the reasoning process and outcomes. We studied the reasoning process by detecting and analyzing faults in the diagnostic reasoning process through retrospective

record reviews. We identified these faults with a low threshold, so they are labeled as suboptimal cognitive acts (SCAs) rather than cognitive errors. Our primary goal was to determine the relationship between SCAs and the occurrence of diagnostic error and patient harm. Furthermore, we examined where in the process (e.g., history taking and physical examination) SCAs occurred. Finally, we determined the causes of SCAs by classifying them according to the model of unsafe acts.¹⁴

Method

We focused our retrospective review on records of admitted patients who were treated for dyspnea at five acute care hospitals in the Netherlands (one university hospital, two tertiary teaching hospitals, and two general hospitals) during a 10-month period in 2007 and 2008.

The Dutch medical education system

In Dutch hospitals, residents generally act as the treating physicians, but residents are supervised by a staff member. The resident typically informs the supervisor about new patients and discusses important decisions with the supervisor.

Please see the end of this article for information about the authors.

Correspondence should be addressed to Ms. Zwaan, EMGO Institute, Department of Public and Occupational Health, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands; telephone: (31) 20-444-9612; e-mail: laura.zwaan@vumc.nl.

Acad Med. 2012;87:149-156.
First published online December 20, 2011
doi: 10.1097/ACM.0b013e31823f71e6

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A71>.

As residency progresses, the residents become more independent, receive more responsibilities, and require less supervision. However, faculty supervisors are always responsible for the residents under their supervision.

Patient selection

The study focused on dyspnea patients, who were admitted to the hospital with dyspnea or became dyspneic during the hospital stay. By selecting dyspnea patients, we selected a homogenous patient group that has not been studied extensively in the field of diagnostic error. Five acute care hospitals in the Netherlands (the VU University Medical Center, of which the EMGO research institute is a part, as well as two tertiary teaching hospitals and two general hospitals, which are not named in order to honor the terms of the institutional consent agreements) participated in the study. The study took place in seven departments of internal medicine, cardiology, and pulmonology. (In some hospitals, the internal medicine department worked closely with one of the other two, and in those cases both participated.) We staggered the initial data-gathering process at each hospital to allow about one month between when we started including patients at one hospital and when we began including patients at the next. Following implementation, every hospital participated for six to eight months between May 2007 and February 2008. The treating physicians in the participating departments included all consecutive dyspnea patients who were admitted to the hospital during the study period at each hospital.

Development of the record review questionnaire

Optimal diagnostic process. Because there was no guideline available in the Dutch handbook for diagnostic guidelines for the diagnostic process when starting with dyspnea as a symptom, we used a two-staged Delphi method to determine the optimal diagnostic process for dyspnea patients. Seven experienced internists participated in this Delphi method between June and December 2006. We selected internists who had recently retired and met the following criteria: They (1) had at least 10 years of postgraduate work experience in internal medicine, (2) had a good reputation amongst colleagues, and (3)

were retired for no longer than five years at the time of selection. Three of the seven internists had been employed at one of the participating hospitals.

The goal of the Delphi study was to determine the steps that had to be considered while diagnosing a dyspnea patient, such as establishing the recent history, and requesting information from previous health care professionals. During a first meeting at the VU University Medical Center, we informed the internists about the research protocol, the goal of the Delphi study, and their role in the process. After they agreed to participate, they received by e-mail 12 cases of real dyspnea patients, which were selected specifically for the Delphi study, and answered open-ended questions regarding the diagnostic process. On the basis of the internists' responses, we developed a first draft of the optimal diagnostic process for dyspnea, including all important diagnostic steps. Next, we invited the internists to a consensus meeting, during which we discussed and adjusted the optimal diagnostic process. The internists received the adjusted version of the optimal diagnostic process, and they commented on this version as well. We integrated their final comments, and all internists approved of the final version (see Supplemental Digital Appendix 1, <http://links.lww.com/ACADMED/A71>).

Review questionnaire. We invited four of the internists who had participated in the Delphi study to participate as record reviewers. They were selected on the basis of their availability and the distance of the commute to the participating hospitals. Three of the four had previously been employed in one of the participating hospitals, so they did not review patient records from those hospitals.

On the basis of the optimal diagnostic process determined in the Delphi study, we developed a first draft of the record review questionnaire to identify SCAs.¹⁵ To refine the questionnaire, we discussed it with the reviewers during a one-day training session using patient records of dyspnea patients. Finally, we adjusted the questionnaire based on the suggestions of the internist reviewers and programmed the questionnaire in an electronic format (Blaise 4.5; Statistics Netherlands; Heerlen, Netherlands). During a subsequent half-day training session, the internist reviewers practiced with the

adjusted review questionnaire and the electronic template. We made additional changes in the questionnaire and the electronic template based on the second training session. We organized a final session to discuss any difficulties the reviewers had experienced during their reviews of patient records. This meeting served two goals: (1) to identify any final necessary changes in the questionnaire, and (2) to increase the interrater reliability by discussing how to address difficult aspects of the questionnaire during the review process.

Record review process

Each eligible patient record was reviewed by one of the four reviewers (a previous study showed that two reviewers did not increase the interrater reliability¹⁶). All patients gave their consent for the study before they were discharged, and the records were reviewed about four months after the patient's discharge. At this time, the records were available and generally included any readmission to the hospital.

The record review focused on identifying SCAs and diagnostic errors. The SCAs were defined as faults in the diagnostic reasoning process, which contributed to or could have contributed to the occurrence of a diagnostic error. A diagnostic error was defined as a diagnosis that was unintentionally delayed, wrong, or missed, as judged from the eventual availability of more definitive information.⁷ The questionnaire assessed all steps of the diagnostic process: (1) history taking, (2) physical examination, (3) laboratory requests, (4) imaging techniques, (5) outlining a diagnosis, (6) starting the treatment, and (7) verification of diagnosis and treatment during the patient's stay. For the first four steps, the review questionnaire inquired whether all relevant aspects (e.g., family history, EKG, D-dimer) were identified or verified by the treating physician. Subsequently, the reviewers determined, on the basis of the optimal diagnostic process and their clinical experience, whether the treating physician interpreted the findings correctly. For steps 5 through 7, the reviewers assessed the treating physician's interpretation of the findings, including whether correct conclusions were drawn from the findings and whether the correct treatment was started based on the symptoms. For any identified SCAs, the reviewers answered

additional questions about their causes and actual or possible patient harm, defined as any disadvantage for the patient that leads to prolonged or strengthened treatment, temporary or permanent impairment, or death.¹⁷ Detailed information on the review process can be found elsewhere.¹⁵ The record review questionnaire is available in Supplemental Digital Appendix 2, <http://links.lww.com/ACADMED/A71>.

Cause classification

The causes for the SCAs were impossible to classify on the basis of the record review alone. Therefore, a researcher (L.Z.) conducted interviews with the treating physicians regarding the SCAs identified during the review process. Of the 72 residents who included patients in the study, 58 were involved with SCAs. We were able to conduct an interview with 48 out of these 58 residents and discussed 144 of the 163 cases with SCAs. The remaining 19 cases (11%) were not discussed because the treating physician could not be contacted for the interview after several attempts via e-mail, telephone, or in person. The cause classification of these cases is, therefore, missing from our analyses.

The interviews involved questions such as “Pulmonary embolism was a possible diagnosis. Why did you decide not to follow-up on the elevated D-dimer?” and “Why did you decide not to perform an EKG to check for heart rhythm disorder?” When several physicians were involved in the diagnostic process of a patient, the interview generally took place with the physician who admitted the patient to the hospital. If applicable, we also interviewed the physician who diagnosed and treated the patient on the ward. The physicians were always interviewed about their own decisions. During the brief interviews, the interviewer took notes and showed them to the residents at the end of the interviews for approval.

We used the interview notes and the record review to classify the causes of each SCA with Reason’s model of unsafe acts,^{14,18} a taxonomy used to classify human errors (see Figure 1 for an explanation of the categories).¹⁹ First, we determined on the basis of the interview whether an SCA was intended or unintended. We defined an intended act to be when the physician acted deliberately but incorrectly, often not

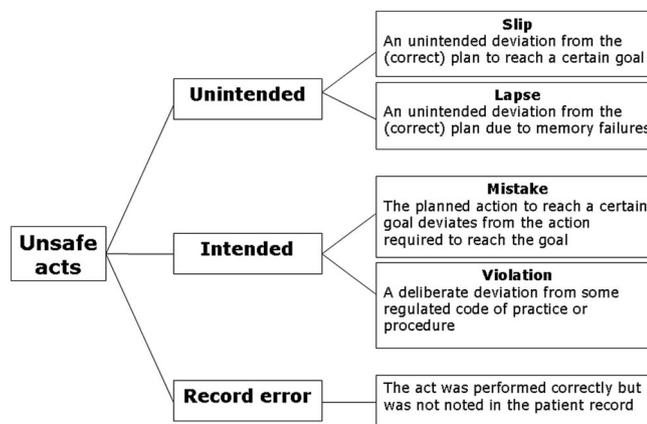


Figure 1 Taxonomy used to classify the causes of suboptimal cognitive acts identified in a record review of 247 dyspnea patient records, 2008, The Netherlands. Adapted from Reason J. Human Error. New York, NY: Cambridge University Press; 1990. Other source material: Reason J. The human factor in medical accidents. In: Vincent CA, ed. Medical Accidents. Oxford, UK: Oxford Medical Publications; 1993.

knowing that the action was incorrect. Unintended acts, on the other hand, were generally deemed to be failures of omission in which the physician intended to do one thing, but, because of something else interfering, the intended act was not performed or was not performed correctly.

Once we determined whether an SCA was intended or unintended, we classified it according to the mutually exclusive categories of the model: mistakes, violations, slips, lapses, or record errors (Table 1). A *mistake* is an intended act, but the physician does not know it is incorrect. For example, the physician did not perform a D-dimer because he or she incorrectly assumed that pulmonary embolism was unlikely. A *violation* is also an intended act, but in this case the physician knows that he or she should have acted differently. For example, the physician decided not to perform a D-dimer to avoid the necessary follow-up tests that would be required if it was elevated. A *lapse* is an unintended act in which the physician intended to do something but forgot to actually perform the task. For example, the physician wanted to perform a D-dimer but forgot to request it on the laboratory form. The second unintended act is a *slip*, in which the plan was correct but was carried out incorrectly. For example, the D-dimer test was not ordered because the physician checked the wrong box. We added the category of “record error” to the model because physicians often indicated that they did perform a certain test or examination but did not note it in the patient record.

After the interview, we discussed the answers of the treating physicians with the internist reviewers to check whether the SCAs found in the record review could still be classified as suboptimal after the explanation given by the treating physician. This was the case for 553 of 565 SCAs (98%). In the few cases in which an SCA was considered invalid on the basis of the physician’s answer, it was generally because of an oversight on the part of the reviewer. For example, the reviewer thought an important test was not conducted, but during the interview it was found among the test results. Furthermore, the context of the occurrence could lead to removing the SCA, such as when a reviewer indicated that an EKG was not performed, but the treating physician clarified that the patient died before the EKG could be conducted.

Reliability study

To assess the interrater agreement, we selected a random sample of 10% ($n = 25$) of the patient records. Each of the 25 records was reviewed independently by any two out of the four possible reviewers. All four internist reviewers were involved in the reliability study. We calculated the interrater reliability for the occurrence of SCAs, diagnostic errors, and patient harm using the Cohen kappa. The reliability for the content of the SCAs could not be calculated because of the unlimited number of possibilities.

Statistical analysis

We used frequency tables to determine the occurrence of SCAs, diagnostic error, and patient harm, the types of causes, patient harm, and possible patient harm. To study

Table 1

Examples of Suboptimal Cognitive Acts (SCAs) Identified During a Review of 247 Dyspnea Patient Records Along With the Treating Physician's Explanation and the Cause Classification of the SCA, 2008, The Netherlands

Diagnostic stage	Description in record review	Explanation by physician	Cause classification
History taking	The physician did not ask about the family history, but the patient had a family history with breast cancer.	"I think it is important to ask the patient about their family history. I forgot to ask the patient in this case."	Lapse
	The physician did not ask the patient whether he had symptoms of aspiration.	"I did not ask for it although it is important. I expected the nurses to obtain this information."	Violation
Physical examination	Central venous pressure was not checked.	"I usually look at the central venous pressure in the neck; however, I do not always write it down when it is not deviating."	Record error
	Pulsus paradoxus was not performed.	"I never do that; to me, it is never an important factor."	Mistake
Laboratory tests	No D-dimer was done, even though, based on the symptoms, pulmonary embolism was not unlikely. (Based on the Wells score, a D-dimer was necessary.)	"I already had a diagnosis for the patient. This was most likely; I did not think the patient had pulmonary embolism. Looking back, maybe I should have checked."	Mistake
	Antibiotics were administered before the sputum test was performed, and therefore the result of the sputum test was not conclusive anymore. However, it was still used as diagnostic information.	"That was wrong. The sputum test was not performed in the emergency department, but later in the ward when the antibiotics were already administered. I just started working as a physician and I did not know that yet."	Mistake
Additional tests and biopsies	EKG was not performed.	"I usually order an EKG for every patient; since it would be relevant for this patient, I probably forgot about it."	Lapse
	High-resolution scan was not performed, though pulmonary embolism was very likely.	"I did not think thrombosis or pulmonary embolism was very likely, and since the patient was already treated with a low dose of fraxiparine, I decided not to make a scan."	Mistake
Diagnosing	Chronic obstructive pulmonary disease (COPD) was not considered during the visit at the emergency department (though present in patient's medical history). The physician focused on heart failure.	"I don't know why COPD was not considered. It seems like I focused on heart failure. That was wrong."	Mistake
	Laboratory results suggested anemia; however, the physician did not notice this.	"I did not notice it."	Slip
Treatment	The patient got a bleeding due to too much oral anticoagulant medication.	"I overlooked the contraindication for oral anticoagulants; this is really bad. Luckily, I found out when the patient was discharged from the hospital, which was after 7 days. The patient did not get any anticoagulants prescribed to use at home."	Slip
Verification	The overload disease was not recognized on time when dyspnea increased during the patient's hospital stay.	"I did not realize the patient was suffering from overload disease, which is why I did not handle as such."	Mistake

the overlap between SCAs, diagnostic error, and patient harm, we used cross tables (see Table 2). We used independent sample *t* tests for comparisons between groups (i.e., to compare means of the number of SCAs in cases with and without diagnostic error, and with and without patient harm).

It was not always possible to determine with certainty whether a diagnostic error occurred (e.g., autopsy reports were lacking). Therefore, we added the cases judged as "probably a diagnostic error" by the expert internists to the category "diagnostic error." The cases in which a diagnostic error and patient harm co-occurred are presented only in the diagnostic error category (Figure 2).

Confidentiality and ethical approval

The medical ethics review committee of the VU University Medical Center approved the research protocol, and all participating hospitals and hospital departments granted approval to participate. The internist reviewers and researchers involved in the data collection signed a confidentiality agreement to maintain the secrecy of the data. Patients included in the study signed informed consent for review of their patient record. The patients were informed that they would not obtain the results of their own record review. The Netherlands does not have a "claim-culture."²⁰ As far as we know, none of the

cases reviewed in this study were involved in malpractice claims.

In case the internist reviewers had concerns about deliberate harmful or illegal acts or negligent behavior, they could discuss them with an ethics committee.

Results

Patient sample

A total of 261 patient records were included in the study, of which 14 records were lost, for example, because of incorrect registration of the location of the patient record by the archive. Expert internists reviewed the patient records of

Table 2

Examples of Types of Cases With Suboptimal Cognitive Acts (SCA), Diagnostic Error, and Patient Harm From 247 Dyspnea Patient Records Reviewed by Four Expert Internists Evaluating Treating Physicians' SCAs, 2008, The Netherlands

Finding in record			No. (%) of records containing mix of findings	Mean number of SCAs per record	Example of case containing mix of findings
SCA	Diagnostic error	Patient harm			
Yes	Yes	Yes	7 (2.8)	7.7	Several factors, such as central venous pressure, low blood pressure, and peripheral pulse deficit were not checked during the physical examination; therefore, heart failure was detected late, which urged additional treatment.
Yes	No	Yes	16 (6.5)	5.2	The patient's medical history was insufficiently taken into account. The patient underwent two radiation sessions for lumbal bone metastases. In addition, the patient had a creatinine of 290. The patient still received postoperative gentamicin, diclofenac, and cefuroxim. The patient's kidneys declined, which contributed to the patient's death.
Yes	Yes	No	21 (8.5)	5.0	No blood and urine cultures on urosepsis were performed, due to which the diagnosis urosepsis was (most probably) missed. However, patient harm was probably prevented by antibiotics for another disease.
Yes	No	No	119 (48.2)	2.7	No EKG was performed while the patient had a deviant heart frequency. The EKG would have been necessary to rule out heart rhythm disorder. In the end, it appeared that the patient did not have a heart rhythm disorder; thus, the patient was not harmed, and no diagnosis was missed.
No	Yes	Yes	1 (0.4)	0	The diagnosis pancreatitis was missed, and therefore the patient unnecessarily suffered from dyspnea. The causes were patient related because the presentation of the disease was atypical.
No	No	Yes	4 (1.6)	0	In response to treatment with fibrinolyse, the patient got a large hematoma and a decrease in hemoglobin, for which a blood transfusion was needed. However, this was not preventable.
No	Yes	No	5 (2.0)	0	Cryptogenic organizing pneumonia (bronchiolitis obliterans organizing pneumonia) was not detected. However, this diagnosis is hard to determine and very rare. The process was correct, and the diagnosis was found a week later and treated correctly.
No	No	No	74 (30.0)	0	Diagnostic process occurred without any suboptimal acts, patient harm, or incorrect diagnosis.

247 patients. The patient characteristics are described in Table 3.

Incidence

A total of 565 SCAs were identified in 163 of the 247 patient records (66%), with an average of 2.3 SCAs per patient record (565 SCAs in 247 records). In 34 of 247 cases (13.8%), a diagnostic error occurred. In 28 of 247 cases (11.3%), the patient was harmed. There were more SCAs in cases with a diagnostic error than in cases without a diagnostic error (mean

of 4.7 versus 1.9 SCAs, respectively; $t = 4.41$, $P < .0001$). Also, there were more SCAs in cases with patient harm as compared with cases without patient harm (4.9 versus 2.0, respectively; $t = 4.11$, $P < .0001$). For an overview of the incidences of the different events and examples of SCAs, see Table 2.

Of the 54 cases with a diagnostic error and/or patient harm, 44 cases (81%) were related to the occurrence of SCAs. However, in 10 of 54 cases with a diagnostic

error and/or patient harm (19%), no SCAs were present (see Table 2).

Overall, SCAs occurred mostly during the early stages of the diagnostic process: history taking, physical examination, request of laboratory tests, and additional tests (see Figure 2).

Causes of SCAs

In total, the causes of 476 SCAs were classified. The SCAs were intended actions in 296 of 476 SCAs (62%),

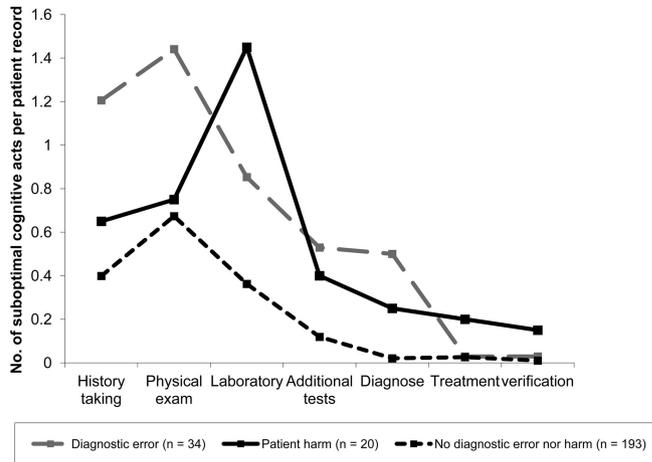


Figure 2 Occurrence of suboptimal acts in the different stages of the diagnostic process for different types of cases as identified in a record review of 247 dyspnea patient records, 2008, The Netherlands.

including 276 (58%) mistakes and 20 (4%) violations. One hundred twenty-four of 476 SCAs (26%) were unintended actions, including 64 (14%) slips and 60 (12%) lapses. Record errors occurred in

56 of 476 (12%) of the SCAs. There were no different cause patterns for the different types of cases (diagnostic error, patient harm, and the cases without those events).

Table 3

Patient Characteristics From 247 Dyspnea Patient Records Reviewed by Four Expert Internists Evaluating Treating Physicians’ Suboptimal Cognitive Acts, 2008, The Netherlands

Characteristic	Results
Male patients, no. (%)	115 (46.6)
Average age (SD)	69.9 (14.9)
Median length of stay (interquartile range)	10 (11)
Final diagnosis, no. (%)	
Heart failure	51 (20.6)
Chronic obstructive pulmonary disease/bronchitis/asthma	59 (23.8)
Pneumonia	47 (19.0)
Malignancy	16 (6.5)
Pulmonary embolism	14 (5.7)
No diagnosis	3 (1.2)
Other (e.g., meningitis, cholecystitis, anemia, septic shock, pneumothorax, viral infection, hyperventilation, hepatorenal syndrome, bronchiolitis, pancreatitis)	57 (23.1)
Relevant comorbidity, * no. (%)	116 (51)
Heart failure	18 (15.5)
Atrial fibrillation	15 (12.9)
Pneumonia	13 (11.2)
Chronic obstructive pulmonary disease/bronchitis/asthma	20 (17.3)
Renal insufficiency	11 (9.5)
Ischemic heart disease without heart failure	11 (9.5)
Malignancy	8 (6.9)
Hypoglycemia	9 (7.8)
Anemia	5 (4.3)
Other (e.g., aorta valve stenosis, meningitis, septic shock, viral infection, lung fibrosis, urinary tract infection, liver failure, HIV, hypertension, pulmonary embolism, thrombosis)	74 (63.8)

* There could be more than one diagnosis per patient.

Patient harm

In 28 out of 247 cases in which the patient was harmed, the most severe consequences were additional interventions or treatment (6 of 28 cases, 21%), health impairment at the time of discharge from the hospital (5 of 28 cases, 18%), readmission to the hospital (4 of 28 cases, 15%), death (2 of 28 cases, 7%), and other types of harm (11 of 28 cases, 39%). The last category included cases in which a late diagnosis led to a late treatment, due to which the patient unnecessarily suffered from symptoms or had an extended hospital stay, as well as cases in which the extent of the harm was unknown at the time of the record review.

Of the SCAs that did not contribute to the occurrence of patient harm, the reasons that the patient was not harmed were identified in 72 cases: In 18 of 72 cases (25%), the correct diagnosis was found before the patient was harmed (e.g., heart failure was detected five days late). In 16 of 72 cases (22%), the treatment for the incorrect diagnosis was also helpful for the missed diagnosis (e.g., oral anticoagulants were prescribed for another medical problem and might have prevented harm elicited by a missed pulmonary embolism). In 11 of 72 cases (15%), the patient was not harmed due to luck (i.e., the patient was not affected by the incorrect treatment). In 6 of 72 cases (9%), the SCAs were too insignificant to lead to patient harm (i.e., unnecessary but not invasive laboratory tests were conducted). In 21 of 72 cases (29%), it was unknown whether the patient was harmed at the time of the record review, for example, because no one followed up on the patient after the patient was discharged.

Reliability

The reliability between the reviewers for the determination of patient harm was fair ($\kappa = 0.34$; 95% CI 0.0–1.0; 88% agreement). For the determination of whether the diagnosis was correct or incorrect, the reliability was moderate ($\kappa = 0.51$; 95% CI 0.06–0.95; 84% agreement).²¹ The reliability for the determination of whether an SCA was present or not was fair ($\kappa = 0.28$; 95% CI 0.00–0.76; 76% agreement).

Discussion

Interpretation

Our results imply that the large variability in the occurrence of diagnostic error found in earlier studies^{5,8,9} is not very surprising. Studies based on autopsy reports only reported diagnostic errors in the cases of deceased patients,⁸ whereas studies on adverse events reported cases in which the patient was harmed because of a diagnostic error.⁵ Our study shows that there is only a small overlap between these two types of outcomes, which is in line with earlier findings.^{5,9,13,22}

Most SCAs occurred during the data-gathering stages because those stages involve asking about and examining many different aspects and are of great predictive value to the correct diagnosis.²³ SCAs occurred in the same diagnostic stages for cases with diagnostic errors as in cases without diagnostic errors. In cases with patient harm, SCAs occurred often during laboratory testing, which involved cases in which unnecessary test results were conducted.

We observed a higher number of SCAs in cases with diagnostic error and patient harm, implying that the number of SCAs during the diagnostic process is more predictive of eventual diagnostic errors and patient harm than are other factors, such as the stage of the diagnostic process in which an SCA occurs or the cause of an SCA. Although it may have been easier to correct a small number of early SCAs during the remaining part of the diagnostic process, it seemed to be more difficult to correct for many SCAs. This might account particularly for mistakes because the treating physicians assumed that they reasoned correctly when, actually, they did not. Another explanation may be that, in cases with more SCAs, it is more likely that one of the SCAs will not be discovered during the diagnostic process, and, therefore, it is more likely to lead to a diagnostic error or to patient harm.

The harm that patients suffered was, in most cases, not severe, which is in contrast with previous studies on diagnostic error.^{1,5} However, in a substantial number of cases there was still the possibility of patient harm, as the final patient outcome was not yet known at the moment of the record review. Also,

even the cases in which the patient was not harmed could have led to patient harm in a different situation. Specifically, the reasons why the patient was not harmed were often indicated to be due to luck or a coincidental correct treatment.

Further research should focus on studying the mechanisms underlying SCAs in an experimental setting. For example, different information-processing strategies (e.g., pattern recognition and hypothetico-deductive reasoning) could be manipulated to see which one leads to the most SCAs. Furthermore, possible interventions should be developed and tested. A first step could be to apply practices already used in aviation research. These interventions have proven to be effective in improving situation awareness and in reducing the number of errors.²⁴ In particular, the strategies to check the validity of one's own assessment may be suitable to apply to the diagnostic reasoning process and may help physicians to become aware of the errors they make and to correct those errors before they result in patient harm.

Strengths and weaknesses

Our study is innovative because it related faults in the diagnostic process to diagnostic errors and patient harm and showed that a correct process does not always lead to a good outcome, and a suboptimal process does not always lead to an adverse outcome. The results are especially valuable because the treating physicians were involved in the initial interpretation of the data, allowing us to determine the causes of the SCAs more precisely. In addition, this study starts with a symptom that has not been studied extensively in the field of diagnostic error and follows the diagnostic process rather than focusing exclusively on patient outcomes. Studying diagnostic error with a focus on the respiratory system broadened the spectrum of types of symptoms and diseases studied in the field of diagnostic error.

Our study has several limitations, however, such as subjectivity in the reviewers' judgment, despite our well-trained panel of experts. Our reliability study shows that there are some differences in judgment of the experts. However, the treating physicians considered their SCAs to be faults also (in

98% of the SCAs), suggesting that the SCAs were correctly identified (high specificity); some SCAs, however, might have been missed (lower sensitivity). A modest reliability is a known problem with patient record reviews,^{5,16} despite various adaptations in the review process to improve the reliability. Furthermore, although we treated the diagnostic process as a step-by-step practice for the purpose of this study, it is actually much more complex than the way it is presented here. Our results may, therefore, not reflect all the difficulties faced by physicians during the diagnostic process.

Another limitation in record review studies is hindsight bias. Looking back, it is difficult not to be influenced by the known outcomes. However, because most of the cases in this study involved SCAs without a diagnostic error and patient harm, we think this bias had little effect on the results of this study. The results may also be affected by physician participants giving socially desirable answers during the interviews. Although the treating physicians were very open and willing to discuss the SCAs, they could have given socially desirable answers to look better or rationalize their answers. This may have affected the cause classifications. Lastly, the records were reviewed shortly after the patients' discharge. It was, therefore, not always possible to determine with certainty whether a diagnostic error occurred and whether or not the patient was harmed, as definitive information about the patient's current health status was not always available.

Conclusions

Faults in the reasoning process are associated with diagnostic error and patient harm. However, adverse outcomes also occurred without any faults in the diagnostic process. Because physicians were often not aware of their SCAs, strategies to prevent faults in the diagnostic process should focus on making physicians aware of their reasoning errors. To achieve this, it may be helpful to study how existing interventions aimed at improving situation awareness, such as self-checking activities, can help improve the diagnostic reasoning process.²⁴ Furthermore, the underlying mechanisms

that lead to the occurrence of SCAs should be studied experimentally.

Ms. Zwaan is a PhD candidate, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands.

Dr. Thijs is internist, Department of Internal Medicine, VU University Medical Center, Amsterdam, The Netherlands.

Dr. Wagner is professor, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, and NIVEL, Netherlands Institute for Health Services Research, Amsterdam, The Netherlands.

Dr. van der Wal is professor, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, and chief, Netherlands Health Inspectorate, Amsterdam, The Netherlands.

Dr. Timmermans is professor, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands.

Acknowledgments: The authors thank the physicians who participated in the Delphi method; Sanne Lubberding, José Jansma, and Inge van Wagtenonk for their help with the data gathering; and the physicians who reviewed the patient records: H.A.P. Asscheman, MD, PhD, W.W. Meijer, MD, K.J. Roozendaal, MD, PhD, and J. Silberbusch, MD, PhD. They would also like to thank the hospitals and their staff for including the patients in the study and for facilitating the patient records.

Funding/Support: This study was supported by the Dutch Society of Medical Specialists, with financial support from the Dutch Ministry of Health, Welfare, and Sport.

Other disclosures: None.

Ethical approval: The medical ethics review committee of the VU University Medical Center reviewed and approved the research protocol.

Previous presentations: The results presented in this manuscript were presented as a poster presentation at the Society for Medical Decision Making and the Diagnostic Error in Medicine Conference, both in Toronto, Ontario, Canada, October 2010.

References

- 1 Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med*. 1991;324:377–384.
- 2 Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care*. 2000;38:261–271.
- 3 Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: The incidence of adverse events among hospital patients in Canada. *CMAJ*. 2004;170:1678–1686.
- 4 Zegers M, De Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: Results of a retrospective patient record review study. *Qual Saf Health Care*. 2009;18:297–302.
- 5 Zwaan L, De Bruijne MC, Wagner C, et al. A record review on the incidence, consequences and causes of diagnostic adverse events. *Arch Intern Med*. 2010;170:1015–1021.
- 6 Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: Current understanding and future directions. *J Clin Oncol*. 2007;25:5009–5018.
- 7 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med*. 2005;165:1493–1499.
- 8 Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. *JAMA*. 2003;289:2849–2856.
- 9 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121:S2–S23.
- 10 Coradazzi AL, Morganti ALC, Montenegro MRG. Discrepancies between clinical diagnoses and autopsy findings. *Braz J Med Biol Res*. 2003;36:385–391.
- 11 Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ*. 2003;37:695–703.
- 12 Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355:2217–2225.
- 13 Neale G, Woloshynowych M, Vincent C. Exploring the causes of adverse events in NHS hospital practice. *J R Soc Med*. 2001;94:322–330.
- 14 Reason J. *Human Error*. New York, NY: Cambridge University Press; 1990.
- 15 Zwaan L, Thijs A, Wagner C, van der Wal G, Timmermans DR. Design of a study on suboptimal cognitive acts in the diagnostic process, the effect on patient outcomes and the influence of workload, fatigue and experience of physicians. *BMC Health Serv Res*. 2009;9:65.
- 16 Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, van der Wal G, de Vet HC. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol*. 2010;63:94–102.
- 17 Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, Waaijman R, van der Wal G. Design of a retrospective patient record study on the occurrence of adverse events among patients in Dutch hospitals. *BMC Health Serv Res*. 2007;7:27.
- 18 Reason J. *The human factor in medical accidents*. In: Vincent CA, ed. *Medical Accidents*. Oxford, UK: Oxford Medical Publications; 1993.
- 19 Reason J. *Human error: Models and management*. *BMJ*. 2000;320:768–770.
- 20 Christiaans-Dingelhoff I, Smits M, Zwaan L, Lubberding S, Van der Wal G, Wagner C. To what extent are adverse events found in patient records reported by patients and healthcare professionals via complaints, claims and incident reports? *BMC Health Serv Res*. 2011;11:49.
- 21 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- 22 Elstein AS. *Clinical reasoning in medicine*. In: Higgs J, Jones MA, eds. *Clinical Reasoning in the Health Professions*. Woburn, Mass: Butterworth-Heinemann; 1995:49–59.
- 23 Schmitt B, Kushner M, Wiener S. The diagnostic usefulness of the history of the patient with dyspnea. *J Gen Intern Med*. 1986;1:386–393.
- 24 Endsley MR, Robertson M. Training for situation awareness. In: Endsley MR, Garland D, eds. *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.