How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews

Thierry Pelaccia, MD, PhD*; Jacques Tardif, PhD; Emmanuel Triby, PhD; Christine Ammirati, MD, PhD; Catherine Bertrand, MD; Valérie Dory, MD, PhD; Bernard Charlin, MD, PhD

*Corresponding Author. E-mail: pelaccia@unistra.fr.

Study objective: The ability to make a diagnosis is a crucial skill in emergency medicine. Little is known about the way emergency physicians reach a diagnosis. This study aims to identify how and when, during the initial patient examination, emergency physicians generate and evaluate diagnostic hypotheses.

Methods: We carried out a qualitative research project based on semistructured interviews with emergency physicians. The interviews concerned management of an emergency situation during routine medical practice. They were associated with viewing the video recording of emergency situations filmed in an “own-point-of-view” perspective.

Results: The emergency physicians generated an average of 5 diagnostic hypotheses. Most of these hypotheses were generated before meeting the patient or within the first 5 minutes of the meeting. The hypotheses were then rank ordered within the context of a verification procedure based on identifying key information. These tasks were usually accomplished without conscious effort. No hypothesis was completely confirmed or refuted until the results of investigations were available.

Conclusion: The generation and rank ordering of diagnostic hypotheses is based on the activation of cognitive processes, enabling expert emergency physicians to process environmental information and link it to past experiences. The physicians seemed to strive to avoid the risk of error by remaining aware of the possibility of alternative hypotheses as long as they did not have the results of investigations. Understanding the diagnostic process used by emergency physicians provides interesting ideas for training residents in a specialty in which the prevalence of reasoning errors leading to incorrect diagnoses is high. [Ann Emerg Med. 2014;64:575-585.]

Please see page 576 for the Editor’s Capsule Summary of this article.

A feedback survey is available with each research article published on the Web at www.annemergmed.com. A podcast for this article is available at www.annemergmed.com.

0196-0644/$-see front matter
Copyright © 2014 by the American College of Emergency Physicians.
http://dx.doi.org/10.1016/j.annemergmed.2014.05.003

SEE EDITORIAL, P. 586.

INTRODUCTION

Background

Clinical reasoning is at the core of medical competence.1-3 It has been the subject of a great deal of research since the 1970s,4 particularly aiming to understand the cognitive activities that lead the physician to make diagnoses.7 Although there are different theories of reasoning, there is a consensus that the context in which a physician is reasoning has a significant influence on reasoning.2,8-10 As several researchers have emphasized,11-12 this observation calls into question the desire to apply these models universally to all situations and environments in which clinical practice takes place, independently of the context and discipline. Emergency medicine is associated with many contextual specifics, eg, the need to act fast, in a context of uncertainty, and to deal with incomplete and partial information while team-managing many patients.15,13-17 These characteristics make the emergency medicine practice environment a complex real-world setting, as defined and studied by naturalistic decisionmaking researchers.18,19 They have led to a recent call for research to identify the particular cognitive strategies used by emergency physicians for reaching a diagnosis.20

Importance

Generating appropriate diagnostic hypotheses and reaching the correct diagnosis are often considered to be the most crucial, complex, and challenging tasks for physicians.20,21 This is particularly the case in emergency practice, in which physicians must make a high number of decisions.22 Furthermore, emergency medicine is one of the specialties in which diagnostic
Editor’s Capsule Summary

What is already known on this topic:
Emergency physicians entertain many diagnoses during each patient encounter. How they are generated and ordered is not well studied.

What question this study addressed
Fifteen experienced emergency physicians, outfitted with face-mounted cameras, provided usual care to a single emergency patient and, while watching the videotape of the encounter, explained their diagnostic hypotheses and the rank order of these hypotheses at key points in the encounter.

What this study adds to our knowledge
Emergency physicians generated three quarters of hypotheses very early in the encounter, one quarter before ever seeing the patient. The hypothesis ranked highest at the outset was the highest and the end in 13 of 15 cases.

How this is relevant to clinical practice
It will not change current practice but can lead to better understanding of emergency physician medical decisionmaking and ways to improve it.

MATERIALS AND METHODS

Study Design
A qualitative approach appeared most suited to the type of research questions we were seeking to answer. Specifically, we used targeted ethnography, a method that consists of collecting data during a limited period on a particular aspect of community activity. This approach is similar to naturalistic decisionmaking research, which uses interviews and field observations to understand how experts actually make decisions in complex, uncertain, and high-stakes real-world environments.

Between May 2011 and April 2012, we carried out semistructured, detailed individual interviews with expert emergency physicians. These interviews were coupled with viewing the video recording of their activity. This technique is considered the most powerful tool in retrospective studies of reasoning in authentic settings, compared with free recall or audio-assisted recall, because the video provides interviewees with rich and vivid cues to explain their thinking during the activity.

There are many issues when using a stand-alone video camera to record someone’s activity. First, it is difficult to record the details of an activity with such a camera. In addition, the presence of an operator is likely to be disruptive. Finally, the camera’s perspective is different from that of the subjects, which
could impair their ability to recall and to explain their reasoning.\textsuperscript{37,38}

As a consequence, we chose to record emergency physicians’ activity from an “own-point-of-view” perspective (Figure 1), using a microcamera mounted at the physician’s eye level. This technique helps participants retrospectively articulate their thought processes by minimizing self-consciousness, by maximizing their psychological immersion in the activity preceding the interview, and by triggering memories of these cognitive processes.\textsuperscript{38-40} This approach, coupled with interviews, has proved to be effective in studying decisionmaking by rugby referees, orienteering runners, and physiotherapists.\textsuperscript{38,39,41} Respondents have been found to seek to render their decisionmaking process explicit and to provide meaning to their actions, rather than to justify or judge them.\textsuperscript{41} Own-point-of-view video-assisted recall also enables subjects to recall more events, to live a greater experiential immersion, and to recollect and describe up to 4 times more detail compared with free recall.\textsuperscript{39} Finally, carrying a camera does not seem disruptive of the processes being studied, and interviewees have no difficulty extracting the relevant information from the video.\textsuperscript{38,39}

The physicians were filmed in their usual work environment while treating a patient admitted to the emergency department (ED) for a potentially life-threatening reason (ie, not a case that could have been dealt with in general practice).

Ethics committee approval for this study was granted by the Education and Social Sciences Research Ethics Committee of the University of Sherbrooke, Québec, Canada (#CER-ESS 2010-71) and by the Committee for the Protection of Persons Northwest 2, Amiens University Hospital, France (#A01586-37).

Selection of Participants

To increase the credibility of our results and support their transferability across sites,\textsuperscript{34} the expert emergency physicians were recruited in 3 different hospitals, none of which was the one in which TP, the emergency physician who conducted the interviews, works: a hospital in the region of Paris, a university hospital in a large city, and a nonuniversity hospital in a medium-sized city. We selected these hospitals because they differed from one another in terms of their geographic location, the density of physicians working in the ED, and the characteristics of the surrounding population. We devised a list of inclusion criteria for “expert” emergency physicians (Figure 2) that deliberately did not take into account years of experience because the validity of this criterion for research purpose has been called into question.\textsuperscript{42-45} In particular, we recruited physicians who had been nominated as “experienced” by their manager.

The physicians were identified by a contact person in each hospital and then met with TP. A consent form stated the study objectives and the fact that participation was voluntary and unpaid. All interviewees signed a written consent form that specifically authorized the video recording of their work activity and the audio recording of the interview. Patients gave oral consent to be videoed.

Data Collection and Processing

Semistructured interviews called “head-mounted video cued-recall interviews” were carried out by TP, with open questions with the video recording as a support to remembering the reasoning involved. The aim of the interviews was to coconstruct results through interactions between experts and a skilled interviewer. The physicians were also asked to quantify (1) their level of certainty relative to the diagnostic hypotheses generated as their examination of the patient continued, using a scale of 0 (diagnostic hypothesis eliminated) to 10 (diagnostic hypothesis confirmed); and (2) the perceived complexity of the case on a scale of 1 (very low level of complexity) to 10 (very high level of complexity). The interviews were held in series of 5 until no new information related to our research objectives emerged from the analysis of the interviews, indicating that the data were saturated.\textsuperscript{39,40} Data collection and analysis took place iteratively.\textsuperscript{34,47} A total of 15 interviews, ie, 5 in each of the 3 hospitals, were carried out several months apart. The median duration of interviews was 53 minutes (interquartile range 40 to 78 minutes).

Primary Data Analysis

We performed a thematic analysis with constant comparison\textsuperscript{47,48} to identify themes that could be used to compare the reasoning of different practitioners and identify commonalities. The interviews were first fully transcribed and then imported into NVivo (QSR International, Melbourne, Australia) qualitative analysis software. This tool was used to facilitate the creation of codes, manual encoding, and storage and recovery of segments of verbatim reports attached to each code. The verbatim reports were subjected to blind primary encoding by TP, CA, and CB, the results of which were discussed between the 3 researchers. Initially, intercoder reliability reached an average of 71% but increased to 96% after discussions. Primary coding followed both a deductive approach, based on a preliminary codebook collaboratively devised by TP, JT, ET, and BC and informed by our theoretical framework, and an inductive approach in which emerging themes were gradually identified and applied to the data. In accordance with recommendations by Miles and Huberman,\textsuperscript{49} the data were then condensed in the form of a matrix for each practitioner to identify the emerging

Inclusion criteria for expert emergency physicians
- have been specifically trained in emergency medicine
- practice emergency medicine full time (reflecting a high volume of activity)
- practice emergency medicine exclusively
- be a hospital emergency medicine specialist
- hold the highest hospital grade (hospital practitioner)
- have been nominated as “experienced” by his/her hierarchical superior

Figure 2. Inclusion criteria for expert emergency physicians.
themes. These matrices were constructed through reading the verbatim reports and primary coding results several times. An interpractitioner comparison matrix based on the identified themes was also produced while switching between the individual matrices and encoded verbatim reports. Discussions between 3 researchers from different disciplines (TP, JT, and ET)—emergency medicine, cognitive psychology, and educational sciences—took place to build the content and organize these matrices. Graphics showing the evolution of the practitioners’ level of certainty relative to each of their hypotheses were then produced.

RESULTS

Characteristics of Study Subjects

Fifteen practitioners were interviewed. The average age of participants was 42 years (SD 5 years). There were 11 men and 4 women, with an average of 12 years’ experience (SD 5 years) in emergency medicine. The median duration of initial patient treatment was 14 minutes (IQR 8.5 to 20.5 minutes).

Main Results

The results are structured according to the specific research questions stated in the introduction.

During the initial patient examination, at which moment did expert emergency physicians generate diagnostic hypotheses?

With the exception of 2 physicians, who had been provided with only a single piece of information (case 7 concerned a patient whom the orderly “had a bad feeling about”; case 13 concerned a patient “in a lot of pain”), all the practitioners generated diagnostic hypotheses before meeting the patient, as soon as they heard or read the initial patient information, generally provided by the triage nurse. Most of the practitioners generated a single hypothesis at this stage, as with physician 4 (Figure 3). More rarely, 2 or 3 hypotheses were formulated during this initial stage of patient treatment, as with physician 8 (Figure 3).

Of all the hypotheses (n=79) generated by the 15 practitioners during initial examination of the patient, a quarter (n=20) were generated before the patient encounter (pre-encounter hypotheses) or within the first 5 minutes of the encounter (early hypotheses), which was generally the first third of the examination. The other hypotheses (late hypotheses) were generated in equal numbers during the second and third thirds of the encounter. In most cases, the level of specificity of these hypotheses was high and referred to a specific pathologic entity. Sometimes, these hypotheses referred to the impairment of a function, a system or an organ, or to a pathophysiologic or causal mechanism.

Overall, no expert generated a single hypothesis. They generated at least 4 and a maximum of 8 diagnostic hypotheses during the initial encounter with the patient. The mean was 5.3 hypotheses (SD 1.9). In two thirds of cases, each hypothesis was generated separately from the others. In a quarter of the situations in which hypotheses were generated, the practitioners interviewed generated 2 diagnostic hypotheses simultaneously, such as physicians 5 and 10 (Figure 3). Three hypotheses were generated simultaneously in 1 case in 10.

How are the diagnostic hypotheses generated?

The hypotheses generated by the physicians before meeting the patient or within the first seconds of this meeting were generated very fast and without conscious effort, as for physician 4 (Figure 3), according to identifying a few clinical or contextual signs provided by a third party or observed in the patient. These hypotheses were sometimes suggested by the third party who passed on the initial patient information.

The following hypotheses were mostly generated during the history and clinical examination, on the basis of identifying information referred to as “key words” by physician 1. The type of cognitive processes used could not always be characterized by

![Figure 3. Participant statements pertaining to diagnostic hypotheses.](image)

![Figure 4. Proportion of hypotheses generated during the different stages of initial patient management.](image)
participants. However, when they did, they usually described an "intuitive" reasoning, as depicted by physician 1, for example, who thought the patient might be in a "postictal state" after experiencing "merging seizures" when he did not identify hemipлегia in the patient: "[I was] rather intuitive at that point, because I'm reasoning from experience...from having already seen this type of patient."

Sometimes new hypotheses emerged from physicians seeking to establish the severity of their initial hypothesis or thinking about a differential diagnosis related to their initial hypothesis. For physician 10, the hypothesis generated this way was the result of conscious effort: "Yes, I'm on the abdomen so I think, OK, I haven't asked other questions, do you have diarrhea, do you have this or that; it's not useful but I'm thinking again while examining her, I think about the GI [gastrointestinal] possibilities... All the GI possibilities are going through my head at that moment, and hey presto! I think of the nausea and say to myself, feels faint, nauseous, so I asked about contraception."

No matter how they were generated, diagnostic hypotheses were immediately attributed a specific level of certainty, especially when several were generated at the same time: [physician 6: about the "renal infection" and "renal colic" hypotheses] "I registered them as possibilities...in the knowledge that the infection was more likely than renal colic because clinically it didn't resemble that. With an infection, patients can be in pain or very unwell. Renal colic, which is like that and doesn't move..., so, speaking scientifically, I assigned [to renal colic] a much lower level of certainty than for a renal infection or even an ulcer."

**How are the diagnostic hypotheses evaluated?**

The practitioners used the hypotheses generated to guide their data collection during the history and clinical examination. All the practitioners said they were looking for information that would "confirm" their hypotheses. This process was described in several ways. Some spoke of a "route," a "track," or a "sequence." Others said it was to "confirm" or "support" (generally for highly probable hypotheses), to "rule out" (generally for unlikely hypotheses), to "orient," "complete," "stick the pieces together," "glue," or "go in that direction, to that corner."

Physicians 1 and 9 clearly talked about the influence of key information on their level of certainty relative to the diagnostic hypotheses evaluated:

- [Physician 1, about the "stroke" and "postictal state" hypotheses] "Here I think I was thinking much more of a postictal state... The fact that there was no real deficit, the fact that it was really a bit fluctuating because I didn’t have the impression that the deficit was worst on the left, so even with preserved movement...I think that the stroke hypothesis was less likely than the postictal state."
- [Physician 9, about the "coronary syndrome" hypothesis, when the patient told him that the chest pain was "continuous"] "I said to myself...it's not really a chest wall pain... I really looked for it but it's really not that. That's funny. Yes, I said to myself, it's not that and he has like a stabbing pain. The fact that he told me it was continuous, that annoyed me a bit because it can be slightly coronary."

During the evaluation of diagnostic hypotheses, most physicians said they remained vigilant to the emergence of alternative hypotheses that they had not yet thought of, which they had "missed" (physician 5) and "forgotten" (physician 3). The aim was to "feed" their pool of hypotheses (physician 6), to avoid reaching a "hasty" conclusion (physician 9), to identify information that "could set them off in a different direction" (physician 1), to "avoid missing something" (physicians 7 and 12), or to "avoid falling into a trap" (physician 11).

Some physicians mentioned the type of processes involved in collecting and interpreting data. They described a "reflex" function, not requiring conscious effort:

- [Physician 7, when the patient mentioned receiving Augmentin] "I don't think—no I'm not thinking—I think that 'Augmentin,' bang! I looked to see if there was an associated rash. That happened automatically."
- [Physician 5, when he roused the patient by shaking him] "He woke up. I didn't even have to think; he looked at me and, well, I was right. My hypothesis [postictal state] was confirmed."

---

**Figure 5.** Generation and rank ordering of diagnostic hypotheses for physician 9. Each mark indicates that a key piece of information had been uncovered that influenced the status of the hypotheses. The horizontal axis shows the chronological process of the initial patient encounter. Time 0:00 marks the instant when the practitioner received the first information about the patient. The vertical axis shows the level of certainty of each hypothesis. It was evaluated by the physicians on a scale of 0 (hypothesis ruled out) to 10 (hypothesis confirmed).
What happens to diagnostic hypotheses at the end of the initial examination?

As indicated in Figures 5 to 7, the diagnostic hypotheses have been rank ordered by the end of the initial encounter with the patient. That means that all participants had one hypothesis with a higher level of certainty than the others. This hypothesis had been generated very early during the patient encounter: in most cases, it had in fact been generated even before the encounter took place; for the other physicians, it had been generated within 90 seconds of the encounter’s beginning. This hypothesis was also considered the most likely at the instant it was generated. With the exception of 2 physicians, it was also the diagnosis selected when the patient left the ED.

Moreover, the physicians interviewed never reached a level of absolute certainty (values 0 or 10 on the graphs) leading them to rule out or formally confirm the hypotheses generated. They said they wanted to wait for the results of investigations (Figure 8). Physician 3, for example, confirmed the hypothesis of fractured pelvis after interpreting the radiographs taken at the patient’s bedside, and physician 1 rejected the hypothesis of hypoglycemia on receiving the blood glucose results.

LIMITATIONS

We collected data from a limited period of patient management. Although highly influential, this early stage is probably not the sole period during which emergency physicians generate and evaluate diagnostic hypotheses. Consistent with our study objectives, we also dissociated action from thinking to focus on the latter, whereas in reality, both occur in parallel and do not follow a linear path. Moreover, although emergency situations were routine and varied, the physicians were confronted with a limited number of them, for which they estimated an average level of complexity of 3.8 of 10 (SD 1.9). Our results therefore cannot necessarily be transferred to all the situations encountered in an ED, particularly in complex cases.

The fact that we tried to quantify several types of data (such as the mean number of hypotheses generated and the periods during which the hypotheses were generated) represents an attempt to provide an overview of the reasoning of our study participants in regard to 1 case each and should not under any circumstances be seen as a desire to generalize our results. The certainty levels used to construct graphs were provided by participants once they were specifically prompted to do so by the interviewer, which could call into question their validity. Although the absolute values of perceived certainty may not be entirely credible, the variations in certainty levels and the rank ordering of hypotheses may provide good valid indications about the way in which participants evaluated hypotheses. Moreover, asking physicians to rank order the lower-priority hypotheses may have yielded an artificial ordering that did not occur while they were actually thinking. We also tried to identify the precise moment when diagnostic hypotheses were generated, but diagnostic impressions may in fact emerge more gradually than suggested by our findings.

Finally, our work relied on the use of retrospective reporting through think-aloud protocols, which are often seen as an

![Figure 6. Generation and rank ordering of diagnostic hypotheses for physician 14.](image)

![Figure 7. Generation and rank ordering of diagnostic hypotheses for physician 15.](image)
Early and intuitive generation of diagnostic hypotheses

The early nature of the generation of diagnostic hypotheses has been well documented since the 1970s.43 It is based on using cognitive processes generally known as "nonanalytical" or "tacit."1,10,54,55 These processes are used automatically and without conscious effort, allowing individuals to make decisions very quickly. In the contemporary, consensual dual-process theory, sometimes presented as a "universal model of diagnostic reasoning," they are referred to as the "intuitive" component (system 1) of information processing.1,55,56

In our study, reasoning was triggered as soon as the practitioners received the first pieces of information about the patient, ie, before the encounter. This enabled them to generate pre-encounter hypotheses based on a largely intuitive procedure. These hypotheses were not always highly specific. This can be explained by the fact that in the emergency setting, the primary goal for a physician is not necessarily to reach a precise diagnosis but to gain a sufficient understanding of the situation to determine his or her next course of action.54

The fact that hypothesis generation is mainly intuitive is consistent with a postulate according to which intuitive reasoning is the individual's default thought mode.55 It is also consistent with naturalistic decisionmaking models such as the recognition-primed decision model. This model postulates that in complex real-world settings, the majority of decisions are made in less than 1 minute on the basis of previous experiences that help decisionmakers quickly match the current situation to patterns they have learned.58 Patterns allow decisionmakers to identify the most relevant cues, determine plausible goals, and promptly implement the most typical course of action, which is crucial in the emergency setting.58,59

More rarely, the generation of diagnostic hypotheses resulted from conscious effort, notably to ensure that no hypothesis had been missed. These results confirm the existence of a second component (system 2) of information processing, known as "analytic" or "deliberate" in dual-process theory.1,56

There are several hypotheses concerning the interactions between system 1 and system 2: both systems could work sequentially, within a context of reasoning that starts intuitively and continues in a form of analytic control intended to refine or modify the hypotheses generated intuitively; system 2 could also be used in atypical clinical situations in which the situation cannot be recognized intuitively.1,10,54,56 In our study, all the physicians reasoned intuitively and only a few hypotheses were generated analytically. Therefore, these results tend to call into question the linear nature of a procedure during which intuitive and analytic reasoning proceed sequentially. According to some naturalistic decisionmaking theorists, intuitive and analytic systems are poles of a continuum (rather than being separated components) that could allow decisionmakers to oscillate from

### DISCUSSION

In this study, we tried to understand how expert emergency physicians make decisions in their complex real-world settings. The interviews concerned real, varied emergency situations, within the normal working environment of the practitioners involved, which is exceptional in clinical reasoning studies, which are often performed out of context.9

Our work revealed that during the initial patient encounter, emergency physicians generated at least 4 and not more than 8 diagnostic hypotheses. Pre-encounter hypotheses were generated even before physicians met the patients. Early hypotheses occurred during the first third of the encounter. A minority of hypotheses—late hypotheses—were generated after this period. Hypotheses were based on the identification of key information and guided data collection that sought to evaluate the likelihood of hypotheses. Usually, these tasks did not require any conscious effort by the practitioners. At the end of the initial patient encounter, the practitioners had one hypothesis in mind that was more probable than the others. They also retained alternative hypotheses until test results could allow them to rule them out.

Figure 8. Participant statements about the need to wait for the results of investigative tests before ruling out or confirming their diagnostic hypotheses. LP, Lumbar puncture.

optimal methodology to capture thought processes.51 A recent study using functional magnetic resonance imaging to investigate functional neuroanatomic differences between thinking versus thinking aloud showed that think-aloud protocol was a reasonable measure of thinking and that it could be used to assess cognition.14 However, this approach is also subject to criticism as a method of exploration of clinical reasoning.52 Specifically, we cannot be certain that the cognitive processes that participants verbalized were the same as those that they actually used while managing the patient.46 A subject's own recall of his or her initial hypotheses may also have been determined by the information he or she received in the rest of the evaluation. We strove to reduce these risks by holding the interviews as soon after the end of the encounter as possible (median 110 minutes; IQR 80 to 180 minutes) and by focusing the interviews on specific aspects of action, thanks to the video recording.
The limited number of diagnostic hypotheses generated

The recognition-primed decision model posits that in complex real-world environments, decisionmakers usually consider a workable solution and do not need to generate a large set of alternatives.59 In fact, generating a large number of hypotheses and seeking to systematically evaluate each would not be desirable because it could lead them to lose control over the situation and to slow decisionmaking.57 This might explain why the practitioners interviewed in our study did not seem to generate all of the potential hypotheses.

The number of hypotheses is also limited because their evaluation relies on the use of working memory, which has limited capacity.62 This could explain why the mean number of hypotheses was 5, which fits with the documented capacity of working memory, defined in several areas of expertise.3,65,66 Time pressure, ie, the need to generate hypotheses in a limited time imposed by the type of job, also affects the efficiency of the process of extracting diagnoses from long-term memory and hence the total number of hypotheses that a person is capable of generating in a given time.55 Time constraints are particularly present in the emergency setting.15,16,20

Finally, the limited capacity of working memory provides explanations about the relatively few hypotheses—the late hypotheses—that were generated after the first third of the meeting with the patient. As working memory gradually fills, it becomes more and more difficult for decisionmakers to generate diagnostic hypotheses. Only hypotheses that reasonably compete with the most likely hypothesis that is being explored can emerge, thus clearly reducing the possibility of generating new hypotheses as the diagnostic process continues.3 The probability that a person will generate a correct hypothesis therefore depends strongly on the propensity to generate it early.53 This observation is reinforced by the fact that the initial hypotheses will guide data collection, and it is therefore highly unlikely that data linked to the correct hypothesis will be considered if the correct hypothesis is not rapidly generated.33 This could also explain why the hypothesis selected on leaving the ED had been generated very early by the practitioners, whether or not this final diagnosis was correct, which we cannot ascertain.

An evaluation of hypotheses based on the identification of key information

The evaluation of diagnostic hypotheses is conventionally carried out within the context of the hypotheticoeductive reasoning model, also called hypothesis verification strategy, documented during the 1970s by Schwartz and Elstein51 and Elstein et al.56,61 This is a very robust concept in the constructionist paradigm.43 According to this model, data collection and interpretation are guided by the physician’s early hypotheses and are used to confirm or rule out these hypotheses.13 In other words, each hypothesis is used to predict clinical findings, were it to be the correct diagnosis. Our results, in a context of emergency medical practice, tend to question 3 principles often associated with this model.

The first concerns the analytic nature of the processes underpinning hypotheticoeductive reasoning.1,3,5,10,55 In fact, we found several cases in which hypothesis evaluation proceeded intuitively rather than analytically, at least in part. The second relates to the linear nature of a model traditionally described as a 2-step sequence: an initial generation of hypotheses followed by hypothesis evaluation.67 Our participants began evaluating their hypotheses as soon as they emerged and continued to generate new hypotheses later on. The third concerns the hypothesis that only complex cases require hypothesis evaluation.26,52 All of the participants in our study evaluated their hypotheses, regardless of case complexity.

Finally, the hypotheticoeductive model might better describe reasoning in static environments in which problems and data do not change all along the decision process,6 unlike what frequently occurs in the emergency medicine practice environment.

We also found that this evaluation process was used to rank order the diagnostic hypotheses. Participant statements about the perceived likelihood of hypotheses were not consistent with bayesian probabilistic reasoning, but rather with more subjective and empirical judgments. These elements tend to confirm naturalistic decisionmaking research that has shown that individuals’ decisions often deviate from the axioms of expected utility or probability theory.18 They are also consistent with the results of a recent study indicating that emergency physicians do not generally reason on the basis of systematic bayesian strategies.58

Diagnostic uncertainty

Concerning the hypothesis evaluation procedure, we also found that participants maintained a certain level of uncertainty during the whole of the initial patient encounter. Uncertainty is one of the main characteristics of emergency medical practice.15,20,21 It is particularly linked to the fact that very common symptoms (eg, chest pain or abdominal pain) can be due to a wide variety of causes.31 It is compounded by the fact that emergency physicians often have only clinical data to use initially.15,16

Maintaining a level of uncertainty and being mindful of the emergence of alternative hypotheses could represent strategies to avoid the most frequent type of diagnostic reasoning error, ie, premature closure.65 This occurs when the physician stops to evaluate alternative hypotheses without having collected enough information because the main hypothesis is perceived as very likely. In our study, no hypotheses were formally confirmed or ruled out without the physicians’ having the results of investigations. The ability to keep an open mind meant they were able to consider the clinical information that emerged as the encounter progressed. Keeping an open frame of mind has recently been identified as an important characteristic of expertise in emergency medicine.6

The implications for training residents in emergency medicine

Residents in emergency medicine need careful supervision because they are quickly required to provide treatment.41 Supervision plays a crucial role in developing learners’ clinical reasoning.31 In this context, feedback is a very important teaching strategy, particularly for developing intuition.26,70-72 Although our study was about experts, the results may help
Generating and Evaluating Diagnostic Hypotheses

Pelaccia et al

Table. Teaching tips to enhance clinical reasoning learning in emergency medicine residents during feedback.

<table>
<thead>
<tr>
<th>How Novice Emergency Residents Make Decisions</th>
<th>What Trainers Should Ask/Check During Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tend to overestimate their level of certainty concerning their initial hypotheses</td>
<td>Probe residents about the likelihood they attribute to their hypotheses</td>
</tr>
<tr>
<td>Have difficulties in rank ordering hypotheses</td>
<td>Check that residents’ data collection strategy is based on hypotheses</td>
</tr>
<tr>
<td>Sometimes use a linear process, extracting the data as they appear, chronologically</td>
<td>Check that residents keep an open mind and consider disconfirming information</td>
</tr>
<tr>
<td>Tend to disregard information that contradicts their main hypothesis and to focus on seeking confirming data</td>
<td>Check that residents generate and investigate more than a single hypothesis (without encouraging them to be exhaustive)</td>
</tr>
<tr>
<td>Use investigations to “fish for information”</td>
<td>Ask residents to justify their prescriptions for tests</td>
</tr>
<tr>
<td>Check whether tests are targeted at hypothesis evaluation rather than hypothesis generation</td>
<td></td>
</tr>
</tbody>
</table>

teachers to target their feedback on the “cognitive gap” that separates novices and experts (Table). In an article published in 2006, Sandhu et al20 wrote: “The 21st-century reality of ED overcrowding, in conjunction with a renewed call to reign in health care spending while minimizing errors in medicine, will force clinicians to search for ever-improving diagnostic and therapeutic efficiency. A better understanding of how decisions are made might open an unrecognized door to these objectives.” Our work provides a better understanding of how emergency physicians reason to reach a diagnosis, a task considered to be both the most crucial and the most complex in an emergency medical environment.

Our findings suggest a wealth of avenues for further research. These include carrying out a more detailed exploration of how the 2 systems, intuitive and analytic, shift from one to the other, and identifying the conditions under which experts are required to reason analytically. All the practitioners interviewed in our study emphasized the remarkable ability of the method used, particularly the use of video footage recorded in an “own-point-of-view” perspective, to help them become conscious of and articulate their reasoning. This approach, which, to our knowledge, has not been used before in an exploration of medical reasoning, could therefore be gainfully used again in further studies. In particular, it could help researchers study errors from a renewed angle that consists in understanding why the physicians’ assessment and actions made sense to them at the time they went wrong, rather than simply identifying when and where they went wrong.

The authors acknowledge all the physicians who took part in the interviews for their invaluable contribution to the research; the physicians who helped us to recruit; Pierre Paillé, PhD, for his contribution to the study design; and Annick Bourget, Erg., PhD, and Nicolas Pelaccia, Engr., for their contribution to this research.

Supervising editor: David L. Schriger, MD, MPH

Author affiliations: From the Centre for Training and Research in Health Sciences Education (CFR-PS), Department of Medical Education, Faculty of Medicine (Pelaccia), and the Faculty of Educational Sciences (Triby), University of Strasbourg, Strasbourg, France; the Prehospital Emergency Care Service (SAMU 67) – Centre for Emergency Care Teaching (CESU 67), Strasbourg University Hospital, Strasbourg, France (Pelaccia); the Department of Pedagogy, Faculty of Education, University of Sherbrooke, Sherbrooke, Québec, Canada (Tardif); the Department of Emergency Medicine, Amiens University Hospital, Amiens, France (Ammirati); the Prehospital Emergency Care Service (SAMU 94), Henri-Mondor Hospital, Public Hospitals of Paris, Créteil, France (Bertrand); the Institute of Health and Society (IRSS), Catholic University of Louvain, Brussels, Belgium (Dory); and the Centre of Pedagogy applied to Health Sciences (CPASS), Faculty of Medicine, University of Montreal, Montréal QC H3T 1J4, Québec, Canada (Charlin).

Author contributions: TP, JT, ET, and BC conceived the study. TP, CA, and CB obtained research funding. TP supervised the conduct of the study and conducted the interviews. TP, CA, and CB coded data. TP, JT, ET, CA, CB, VD, and BC were involved in interpreting the study findings. TP drafted the article, and all authors disclosed any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist and provided the following details: This study was supported by a grant from the French Society for Emergency Medicine.


The study sponsor had no role in the study design, data collection, analysis and interpretation, and decision to write and submit this article. The researchers are independent from the sponsor.

REFERENCES


Volume 64, No. 6 : December 2014  Annals of Emergency Medicine  583
Generating and Evaluating Diagnostic Hypotheses

56. Croskerry P. Context is everything or how could I have been that stupid? Healthc Q. 2009;12:e171-e176.