

# Clinical Reasoning Tasks and Resident Physicians: What Do They Reason About?

Elexis McBee, DO, MPH, Temple Ratcliffe, MD, Mark Goldszmidt, MD, MHPE, Lambert Schuwirth, MD, PhD, Katherine Picho, PhD, Anthony R. Artino Jr, PhD, Jennifer Masel, MD, and Steven J. Durning, MD, PhD

## Abstract

### Purpose

A framework of clinical reasoning tasks thought to occur in a clinical encounter was recently developed. It proposes that diagnostic and therapeutic reasoning comprise 24 tasks. The authors of this current study used this framework to investigate what internal medicine residents reason about when they approach straightforward clinical cases.

### Method

Participants viewed three video-recorded clinical encounters portraying common diagnoses. After each video, participants completed a post encounter form and

think-aloud protocol. Two authors analyzed transcripts from the think-aloud protocols using a constant comparative approach. They conducted iterative coding of the utterances, classifying each according to the framework of clinical reasoning tasks. They evaluated the type, number, and sequence of tasks the residents used.

### Results

Ten residents participated in the study in 2013–2014. Across all three cases, the residents employed 14 clinical reasoning tasks. Nearly all coded tasks were associated with framing the encounter or diagnosis. The order in which residents used specific tasks varied.

The average number of tasks used per case was as follows: Case 1, 4.4 (range 1–10); Case 2, 4.6 (range 1–6); and Case 3, 4.7 (range 1–7). The residents used some tasks repeatedly; the average number of task utterances was 11.6, 13.2, and 14.7 for, respectively, Case 1, 2, and 3.

### Conclusions

Results suggest that the use of clinical reasoning tasks occurs in a varied, not sequential, process. The authors provide suggestions for strengthening the framework to more fully encompass the spectrum of reasoning tasks that occur in residents' clinical encounters.

**B**roadly defined, clinical reasoning is the process “that enables practitioners to take wise action, meaning to take the best justified action in a specific context.”<sup>1</sup> Given this definition, clinical reasoning can be further divided into diagnostic reasoning and therapeutic reasoning. Diagnostic reasoning refers to how a clinician establishes a diagnosis, and therapeutic reasoning refers to how a clinician decides on a plan of action.<sup>2,3</sup>

Previous research has revealed information about factors that should influence management decisions, but this research provides less insight into how physicians actually weigh multiple medical, social, or psychological factors to develop a plan of action.<sup>4,5</sup> Taken together, extant research provides an incomplete picture of how the processes of diagnostic and therapeutic reasoning occur and what information and actions

practitioners must consider to achieve a successful clinical outcome.

Recently, Goldszmidt and colleagues<sup>6</sup> developed a unified list, or framework, of clinical reasoning tasks and proposed that diagnostic and therapeutic reasoning are likely not simple constructs but instead complex processes comprising multiple tasks. In their article, the authors outlined 24 clinical reasoning tasks (List 1), as agreed on by expert opinion, that are thought to occur during clinical encounters. These tasks describe what physicians may do and consider during an encounter, and collectively they provide a unique framework through which to consider what clinicians reason about.<sup>6</sup> The authors did not attempt to define how clinicians may use each of the tasks, the frequency of use, or the sequence of typical use in a clinical setting, nor did the authors identify the extent to which each task should occur in any given encounter. Thus, research is needed to determine if and how physicians use these clinical reasoning tasks in an encounter. In addition, questions remain about whether any variability in how clinicians use these tasks, how the context of the clinical encounter or the expertise of the physician influences use, and if these

factors (variability, context, expertise) impact diagnostic or therapeutic reasoning accuracy.

Ecological psychology is a theory that can be applied to clinical reasoning. It argues that human actions are the result of affordances (recognized opportunities for actions) and effectivities (abilities for action).<sup>7,8</sup> These affordance–effectivity dyads are interdependent. This interdependence is based, in part, on the specifics of a situation.<sup>7,8</sup> For example, a young child can recognize the affordance of a pool (having fun), but he may not be able to swim (does not have the effectivity). Add waves to the pool and that leads to a different set of affordances and effectivities. Likewise, a resident physician's performance in an encounter may be based on various affordances and effectivities that likely vary according to the specifics of the situation; that is, each resident physician may use different reasoning tasks which could result in different diagnostic and/or therapeutic decisions.

In the context of clinical reasoning, ecological psychology would predict that multiple resident physicians would not likely arrive at a diagnosis or

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

Correspondence should be addressed to Elexis McBee, Naval Medical Center San Diego, 34800 Bob Wilson Dr., San Diego, CA 92134; telephone: (619) 532-5288; e-mail: [elexis.c.mcbee@gmail.com](mailto:elexis.c.mcbee@gmail.com).

*Acad Med.* 2016;91:1022–1028.

First published online December 8, 2015  
doi: 10.1097/ACM.0000000000001024

## List 1

**List of Clinical Reasoning Tasks Proposed by Goldszmidt and Colleagues in 2013<sup>a</sup>***Framing the encounter*

1. Identify active issues
2. Assess priorities (based on issues identified, urgency, stability, patient preference, referral question, etc.)
3. Reprioritize (based on assessment, patient perspective, unexpected findings, etc.)

*Diagnosis*

4. Consider alternative diagnoses and underlying cause(s)
5. Identify precipitants or triggers to the current problem(s)
6. Select diagnostic investigations
7. Determine the most likely diagnosis and underlying cause(s)
8. Identify modifiable risk factors
9. Identify complications associated with the diagnosis, diagnostic investigations, or treatment
10. Assess rate of progression and estimate prognosis
11. Explore physical and psychosocial consequences of the current medical conditions or treatment

*Management*

12. Establish goals of care (treating symptoms, improving function, altering prognosis or cure; taking into account patient preferences, perspectives, and understanding)
13. Explore the interplay between psychosocial context and management
14. Consider the impact of comorbid illnesses on management
15. Consider the consequences of management on comorbid illnesses
16. Weigh alternative treatment options (taking into account patient preferences)
17. Consider the implications of available resources (office, hospital, community, and inter- and intraprofessionals) on diagnostic or management choices
18. Establish management plans (taking into account goals of care, clinical guidelines/evidence, symptoms, underlying cause, complications, and community spread)
19. Select education and counseling approach for patient and family (taking into account patient's and family's level of understanding)
20. Explore collaborative roles for patient and family
21. Determine follow-up and consultation strategies (taking into account urgency, how pending investigations/results will be handled)
22. Determine what to document and who should receive the documentation

*Self-reflection*

23. Identify knowledge gaps and establish a personal learning plan
24. Consider cognitive and personal biases that may influence reasoning

<sup>a</sup>Reproduced with permission from Goldszmidt et al.<sup>6</sup>

therapy using the same affordances and effectivities, nor would the sequence of the affordances and effectivities be the same across clinical cases because each diagnostic situation would likely provide unique affordances and effectivities to each individual physician. This unique confluence of a clinical case and physician could relate to the unified list of clinical reasoning tasks,<sup>6</sup> and thus we undertook this study—using the lens of ecological psychology—to further explore tasks related to diagnostic and therapeutic decision making in resident physicians.

The aim of this study, which was part of a larger research program investigating clinical reasoning, was to evaluate the

verbalized clinical reasoning processes of resident physicians in internal medicine. Our aim was to describe the following: (1) what clinical reasoning tasks occurred during residents' clinical encounters, (2) with what frequency residents used different types of clinical reasoning tasks, (3) whether the use of clinical reasoning tasks occurred in a sequential or nonlinear manner, and (4) whether any types of clinical reasoning tasks occurred that are not accounted for by Goldszmidt and colleagues' framework.<sup>6</sup> The number and variety of tasks used may actually be associated with why teaching and assessing clinical reasoning is difficult and could provide insight into why context

specificity (variability in performance within the same content area) occurs.

**Method****Participants**

We invited all resident physicians in internal medicine training programs in the National Capital and the San Antonio Uniformed Services Health Education Consortiums to take part in this mixed-methods study (2013–2014). A research assistant invited the residents to participate through e-mail. We developed no criteria for excluding residents. We obtained informed consent from each resident prior to the study, and we offered no incentives for participating. The institutional review board (IRB) at the Uniformed Services University of the Health Sciences approved the study, and then the Brooke Army Medical Center IRB acknowledged and also approved the study in a memo.

**Design**

This study design, which we have used previously, is part of a program of research whose methods have been previously published.<sup>3,9</sup> Participants viewed three video recordings, each lasting about three to five minutes. Then, they each completed a computerized, free-text, postencounter form. Next, the participants completed a think-aloud protocol while viewing the video recordings a second time. The videotapes portrayed one of three cases, each with a particular diagnosis and each featuring specific contextual factors:

- Case 1, HIV in a patient for whom English is a second language;
- Case 2, colorectal cancer in a patient who is presenting with emotional volatility; and
- Case 3, type 2 diabetes mellitus in a patient with a combination of both low English proficiency and emotional volatility.

The residents watched the videos in a random order.

A group of six medical education experts in the field of internal medicine wrote the scripts for the three videos. The experts intended for the cases to be of equal intrinsic difficulty and for them to present, collectively, a variety of selected contextual factors. These experts also determined what were correct, partially

correct, or incorrect diagnostic and therapeutic response options based on the clinical information presented in each video recording. The video recordings, each featuring a single patient and a single practicing physician, were filmed under the guidance of a study investigator (S.J.D.), and professional actors played the role of patient and physician. In the final stage of production, the group of medical education experts reviewed the cases for authentic portrayal. We have used the videos in a prior experiment, and faculty have demonstrated variation in performance<sup>3</sup> suggesting that the videos might be optimal for studying the use of clinical reasoning tasks in residents.

After watching each video-recorded clinical encounter, the participants completed a post encounter form so that we could assess their clinical reasoning.<sup>10</sup> The form, which had been previously validated, asks participants to identify the following items: (1) what additional history or physical exam information they would seek, (2) what their differential diagnosis would be, (3) what their diagnosis was and which data support that diagnosis, and (4) what diagnostic and treatment plan they would institute for the patient. Immediately after completing this form, the participants rewatched the videos while engaging in a think-aloud protocol facilitated by a research assistant. The instructions for the think-aloud protocol were to arrive at, minimally, a diagnosis for the case. Participants were asked to “think aloud” if they did not speak for five or more seconds. The think-aloud protocol<sup>11</sup> is a standard technique in which participants are asked to verbalize their thinking. When conducted properly, think-aloud protocols are a trustworthy method for capturing the thought processes of research participants.<sup>12–14</sup> The videos provided a standard stimulus on which the participants reflected. We used the think-aloud protocol in an effort not to constrain or significantly influence the thought process. For the same reasons, we gave participants as much time as they needed to complete the think-aloud protocol.

### Data analysis

We coded the qualitative data from the think-aloud protocol using the framework for clinical reasoning tasks published by Goldszmidt and colleagues<sup>6</sup> (List 1). Specifically, two of us (E.M., T.R.) used a constant comparative

approach to individually conduct iterative coding of utterances and to classify the utterances by task number. Following the initial coding, two of us (E.M., T.R.) met to discuss the reasoning tasks and to resolve any disparities. We (E.M., T.R.) also coded all transcripts for number of differential diagnoses, diagnostic accuracy, and presence of diagnostic uncertainty. We coded this information in an effort to determine whether the use of certain reasoning tasks relates to accuracy of diagnostic reasoning. We coded accuracy of final diagnosis as either correct or incorrect, and we defined the presence of diagnostic uncertainty as the inability of the resident to commit to a final diagnosis. As with the classifying of utterances, two of us (E.M., T.R.) coded the information on diagnoses individually and then met to resolve differences by discussion. Using SPSS 22.0 (IBM, Armonk, New York), we calculated descriptive statistics, including means, standard deviations (SDs), modes, ranges, and proportions for accuracy of diagnosis, presence of diagnostic uncertainty, the average number of differential diagnoses, types of tasks used, number of times residents used tasks, and the sequence in which residents used tasks. We (E.M., T.R.) met for a final time to review coding and to resolve differences by consensus.

### Results

A total of 10 residents (4 males, 6 females) participated in the study. Three residents were in postgraduate year (PGY)-1, 3 residents in PGY-2, and 4 residents in PGY-3 of training. Agreement between the two of us (E.M., T.R.) who coded the think-aloud transcripts was greater than 90%, and we reached complete agreement (100%) through consensus.

Across all three cases, the 10 residents employed 14 clinical reasoning tasks. Table 1 displays the mean, mode, and range for each clinical reasoning task used by individual residents in each case, and the number of residents using each task per case. For *Case 1*, the 10 residents used 12 different types of reasoning tasks, each individual resident verbalized an average of 4.4 tasks (range 1–10) during the think-aloud protocol, and the average number of total task utterances (including repeat use of tasks) per participant was 11.6 (range 1–18).

For *Case 2*, the 10 residents used 10 different types of reasoning tasks, each resident verbalized an average of 4.6 tasks (range 1–6) during the think-aloud protocol, and the average number of task utterances (including repeat use of tasks) per participant was 13.2 (range 1–24). For *Case 3*, the 10 resident participants used 11 different types of reasoning tasks, each participant verbalized an average of 4.7 tasks (range 1–7) during the think-aloud protocol, and the average number of task utterances (including repeat use of tasks) per participant was 14.7 (range 1–26).

The order in which participants verbalized the reasoning tasks was highly variable. Across all cases, the residents verbalized Task 1 (Identify active issues) first in 17 (57%) of the think-aloud transcripts and Task 4 (Consider alternative diagnoses and underlying cause[s]) in 10 (33%) of the transcripts. Interestingly, the residents frequently verbalized those two tasks second, if not first. They verbalized Task 4 second in 18 (60%) of the transcripts and Task 1 second in 5 (17%) of transcripts. Task 8 (Identify modifiable risk factors) was verbalized third in 10 (33%), and Task 7 (Determine the most likely diagnosis and underlying cause[s]) third in 9 (30%), of the think-aloud transcripts. The task the residents most frequently verbalized was Task 4; the 10 residents mentioned this task a total of 26 times across the three cases. Only one participant verbalized reasoning tasks related to management in any of the cases, using Task 12 (Establish goals of care), Task 18 (Establish management plans), and Task 21 (Determine follow-up and consultation strategies) more than once and in multiple cases. One participant verbalized self-reflection Task 23 (Identify knowledge gaps and develop a learning plan) a single time. No other participants verbalized self-reflection tasks. To demonstrate how we investigators interpreted spoken utterances, we have provided example utterances for each task in Table 2.

We observed the residents verbalizing several reasoning tasks that were not adequately captured by the clinical reasoning framework. First, the residents verbalized a restructuring or reprioritization of their differential diagnosis in 12 (40%) of the manuscripts, and at least once for all three cases.

Table 1

**Clinical Reasoning Tasks Used by Internal Medicine Resident Physicians in Think-Aloud Protocols for Each Clinical Case, 2013 to 2014**

Clinical reasoning task	Case 1 (HIV)				Case 2 (colorectal cancer)				Case 3 (diabetes mellitus)			
	N <sup>a</sup>	Mean <sup>b</sup>	Mode <sup>b</sup>	Range <sup>c</sup>	N <sup>a</sup>	Mean <sup>b</sup>	Mode <sup>b</sup>	Range <sup>c</sup>	N <sup>a</sup>	Mean <sup>b</sup>	Mode <sup>b</sup>	Range <sup>c</sup>
<b>Framing the encounter</b>												
1	6	1.1	0	0–3	8	1.8	4	0–4	7	1.6	4	0–5
2	2	0.2	0	0–1	0	—	—	—	1	0.1	1	0–1
3	4	0.6	0	0–2	3	0.4	2	0–2	0	—	—	—
<b>Diagnosis</b>												
4	9	5.4	8	0–9	9	5.3	11	0–12	8	5.9	13	0–14
5	5	1.1	0	0–3	0	—	—	—	2	0.2	1	0–1
6	4	0.8	0	0–4	5	1.6	5	0–6	4	0.8	3	0–5
7	5	0.9	0	0–4	6	1.3	4	0–5	8	2.6	6	0–5
8	6	1.2	0	0–5	6	1.0	3	0–5	2	0.3	2	0–2
9	0	—	—	—	1	0.2	2	0–2	7	1.2	3	0–3
10	0	—	—	—	0	—	—	—	0	—	—	—
11	0	—	—	—	1	0.1	1	0–1	1	0.1	1	0–1
<b>Management</b>												
12	1	0.1	0	0–1	1	0.1	0	0–1	0	—	—	—
13	0	—	—	—	0	—	—	—	0	—	—	—
14	0	—	—	—	0	—	—	—	0	—	—	—
15	0	—	—	—	0	—	—	—	0	—	—	—
16	0	—	—	—	0	—	—	—	0	—	—	—
17	0	—	—	—	0	—	—	—	0	—	—	—
18	1	0.1	0	0–1	0	—	—	—	1	0.3	0	0–3
19	0	—	—	—	0	—	—	—	0	—	—	—
20	0	—	—	—	0	—	—	—	0	—	—	—
21	1	0.1	0	0–1	1	0.1	0	0–1	1	0.1	0	0–1
22	0	—	—	—	0	—	—	—	0	—	—	—
<b>Self-reflection</b>												
23	1	0.1	0	0–1	0	—	—	—	0	—	—	—
24	0	—	—	—	0	—	—	—	0	—	—	—

<sup>a</sup>N indicates the number of participants using the code for the specific case; maximum N = 10.

<sup>b</sup>Mean and mode for the number of times participants used the code in the think-aloud protocol for each case.

<sup>c</sup>Range of the number of times each participant used the code in the think-aloud protocol for each case.

Second, residents identified the presence of nonmodifiable risk factors that could not be accounted for by Task 8 (Identify modifiable risk factors) in 11 (37%) of the manuscripts and, again, at least once for all three cases. Third, residents commented on how the use of prior therapies influenced their consideration of the differential diagnosis in 2 (7%) of the cases. Example utterances for each of these unclassified tasks are also listed in Table 2.

As displayed in Table 3, the proportion of residents able to make a correct diagnosis was 30% ( $n = 3$ ), 60% ( $n = 6$ ), and 80% ( $n = 8$ ) for, respectively, Case 1, 2, and 3. The proportion of residents verbalizing diagnostic uncertainty was 90%

( $n = 9$ ), 50% ( $n = 5$ ), and 40% ( $n = 4$ ) for, respectively, Case 1, 2, and 3. The average number of differential diagnoses generated for Case 1 was 4 ( $SD = 2.0$ ); for Case 2, 6 ( $SD = 3.2$ ); and for Case 3, 6 ( $SD = 3.0$ ).

## Discussion

This study explored the clinical reasoning tasks that residents verbalized through a think-aloud protocol while viewing straightforward cases in internal medicine. Participants employed a number of different reasoning tasks. The number and variety of tasks employed is consistent with ecological psychology; that is, use of reasoning tasks varies depending

on unique affordances and effectivities in each clinical situation. Of the 24 reasoning tasks proposed by Goldszmidt and colleagues,<sup>6</sup> the participants, in aggregate, verbalized 14 different tasks across all three cases, and the average ranged from 4.4 to 4.7 tasks per case. The most frequently verbalized reasoning tasks were related to diagnosis and framing the encounter. Specifically, tasks related to considering a differential diagnosis, identifying active issues, establishing a lead diagnosis, selecting diagnostic investigations, and identifying modifiable risk factors were used most often.

The order in which residents verbalized reasoning tasks was highly variable both



among participants and cases. This variability is consistent with ecological psychology; each dyadic interaction of individual and environment produces

a unique encounter. In general, early in the encounter, participants' verbalized tasks related to identifying active issues or establishing a differential diagnosis. As

the encounter progressed, participants' tasks transitioned more to refining the differential diagnosis, considering a lead diagnosis, and selecting diagnostic investigations to help confirm the lead diagnosis. If the residents verbalized management and self-reflection tasks, they did so toward the middle or end of the encounter. Overall, across all 30 manuscripts, the trajectory of tasks appeared to occur in a purposeful but nonlinear and varied manner. This finding could relate to context specificity or the idea that performance during any given clinical encounter is only weakly prognostic of the same individual's performance during the next clinical encounter.<sup>3,15,16</sup> In other words, consistent with ecological psychology, participants in this study demonstrated substantial variability in how they carried out clinical reasoning tasks across cases, suggesting that the clinical reasoning processes needed for Case 1 were not necessarily the same processes needed for Case 2 or 3.

Another tendency that we observed was for residents to use reasoning tasks repeatedly throughout the encounter—even though the actual number of tasks they used was somewhat limited. The repeated use of these specific tasks appeared to proceed in a nonlinear but logical and purposeful fashion. For example, one resident used the following tasks during Case 1: Task 1 three times, followed by Task 7, Task 8 twice, Task 3, Task 4, Task 3, Task 4, Task 5, Task 4, Task 5, and Task 4 five times before finally reaching a final diagnosis (Task 7). Ecological psychology supports this tendency because people's affordances and effectivities may differ, but only within a range of acceptable possibilities for the successful completion of a particular task.<sup>8</sup> Although, in the task of diagnosing a patient with a common complaint in internal medicine, several potential "trajectories" to a successful outcome are possible, these options are bounded. This range of options may offer a potential explanation for why reasoning tasks within each category occurred repeatedly but not in a sequential fashion or in totality.

Notably, the verbalization of management or self-reflection tasks among our research participants was quite limited: Only one participant verbalized any management task, and only one participant verbalized a single self-

Table 2

### Example Utterances of Clinical Reasoning Tasks Used by Resident Physicians in Think-Aloud Transcripts, 2013 to 2014

Reasoning task <sup>a</sup>	Example
<b>Framing the encounter</b>	
1. Identify active issues	"Her chief complaint is losing weight with a good appetite, associated with gastrointestinal symptoms to include bloating and constipation." (Case 2: Colorectal cancer)
2. Assess priorities	"Very important question to try and find out, the chronicity of this." (Case 3: Diabetes mellitus)
3. Reprioritize (based on assessment patient perspective, unexpected findings, etc.)	"Now that is very alarming, that she had to buy new clothes." (Case 2: Colorectal cancer)
<b>Diagnosis</b>	
4. Consider alternative diagnoses and underlying cause(s)	"You also think of type 1 diabetes mellitus, not just type 2." (Case 3: Diabetes mellitus)
5. Identify precipitants or triggers to the current problem(s)	"Somebody around her might be sick ... maybe she has a sick contact." (Case 1: HIV)
6. Select diagnostic investigations	"Evaluation for potential [gastrointestinal] malignancy, start with a colonoscopy with anemia workup." (Case 2: Colorectal cancer)
7. Determine the most likely diagnosis and underlying cause(s)	"She came in for the increased drinking of water and I think it's due to new-onset diabetes." (Case 3: Diabetes mellitus)
8. Identify modifiable risk factors	"Did he use protection (during sexual intercourse)?" (Case 1: HIV)
9. Identify complications associated with the diagnosis, diagnostic investigations, or treatment	"Her decreased sensation to light touch on her soles suggests that she has peripheral neuropathy, probably related to long-standing type 2 diabetes." (Case 3: Diabetes mellitus)
11. Explore physical and psychosocial consequences of the current medical conditions or treatment	"She also has had a candidiasis infection, which suggests that she probably has high glucose levels." (Case 3: Diabetes mellitus)
<b>Management</b>	
12. Establish goals of care	"Going through all her concerns, would start with symptomatic treatment of her constipation." (Case 2: Colorectal cancer)
18. Establish management plans	"Once confirmed, I would treat her for diabetes (mellitus) with first agent metformin." (Case 3: Diabetes mellitus)
21. Determine follow-up and consultation strategies	"I would have close follow-up with her next week. If it has not resolved, look at other things such as [complete blood count], stool samples, swab for strep (pharyngitis) or mononucleosis." (Case 1: HIV)
<b>Self-reflection</b>	
23. Identify knowledge gaps and establish a personal learning plan	"I'm not personally very good at differentiating between the two ... strep (pharyngitis) over mononucleosis." (Case 1: HIV)
<b>Reasoning tasks not captured by the current framework</b>	
Restructure and reprioritize the differential diagnosis	"I really want to get more sexual history and am now concerned for hepatitis or primary HIV infection." (Case 1: HIV)
Identify nonmodifiable risk factors	"She has a family history of colon cancer ... which could be a risk factor." (Case 2: Colorectal cancer)
Consider the impact of prior therapies	"Tylenol is helping her with the pain but it is not specific." (Case 1: HIV)

<sup>a</sup>Clinical reasoning tasks that the participants did not use are not shown.

Table 3

**Number of Residents Identifying the Correct Diagnosis and Number Demonstrating Diagnostic Uncertainty by Case in an Evaluation of the Use of Clinical Reasoning Tasks in Resident Physicians, 2013 to 2014**

Case	No. (% of 30) residents
<b>Case 1: HIV</b>	
Identified correct diagnosis	3 (30)
Verbalized diagnostic uncertainty	9 (90)
<b>Case 2: Colorectal cancer</b>	
Identified correct diagnosis	6 (60)
Verbalized diagnostic uncertainty	5 (50)
<b>Case 3: Diabetes mellitus</b>	
Identified correct diagnosis	8 (80)
Verbalized diagnostic uncertainty	4 (40)

reflection task. One potential explanation for the limited use of management or self-reflection tasks was the high level of diagnostic uncertainty present across participants (Table 3). The presence of diagnostic uncertainty indicates decreased confidence in the leading diagnosis and may have impacted the likelihood that participants would articulate a therapeutic plan for the patient. Previous research<sup>3,9</sup> suggests that modification of contextual factors in the encounter can increase cognitive load. Because of potential increases in cognitive load, participants may have missed key information that would allow for closure of diagnostic reasoning and subsequent development of an actionable therapeutic plan. As an advanced skill, self-reflection may have also been negatively impacted by the presence of increased cognitive load. The increasing number or complexity of contextual factors and its effect on cognitive load—and, in turn, on the use of reasoning tasks—warrants further investigation. Alternatively, the very small number of verbalized therapeutic reasoning tasks may reflect the think-aloud instructions, which asked participants to arrive at, minimally, a diagnosis. Participants may not have verbalized therapeutic reasoning tasks because they had already met minimum expectations.

We observed several reasoning behaviors that were not captured adequately

by the current proposed framework of clinical reasoning tasks. First, participants frequently—in 12 (40%) of 30 transcripts—verbalized a restructuring or reprioritization of their differential diagnosis, while remaining uncommitted to a leading diagnosis. Neither Task 4 (Consider alternative diagnoses and underlying cause[s]) nor Task 7 (Determine most likely diagnosis and underlying cause[s]) seems to accurately capture this task.

Second, Task 8 relates to identification of modifiable risk factors. We propose that this be expanded to also include *nonmodifiable* risk factors. In 11 instances (37% of transcripts), our participants mentioned nonmodifiable risk factors as part of their diagnostic reasoning. For example, a positive family history of colorectal cancer could be a nonmodifiable risk factor for that disease.

Third, we propose that Task 3 (Reprioritize based on assessment) be expanded to include assessment of how the use of prior therapies or interventions may or may not alter clinical reasoning processes. For example, worsening symptoms, in the case of a patient with chronic low-back pain who has already had a six-week course of physical therapy, would have the potential to change both how the encounter is framed and the process of diagnostic reasoning.

We recognize that there are limitations to this study. First, we conducted this study with only a small number of participants ( $n = 10$ ) across three PGY levels. However, we reached saturation of the coding of tasks midway through coding the transcripts, which indicates that our sample was adequate. Second, because only residents were sampled, results may differ in experts who have more advanced illness scripts. We might, for example, hypothesize that experts use fewer tasks (more pattern recognition) for the purpose of making the primary diagnosis but use more tasks related to the nuances of both diagnosis and treatment because of their broader set of affordances and effectivities for practice in their field. Third, all participants were from internal medicine programs, limiting generalizability to other specialties and subspecialties. Fourth, because the participants watched *video-recorded* portrayals of clinical scenarios and *then* responded to a think-aloud

protocol, their verbalized reasoning tasks may not be transferable to how and in what order clinical reasoning tasks are used during real clinical encounters. In addition, this order presents the risk of hindsight; however, we believe that the immediacy between completing the post encounter form and rewatching the video while participating in the think-aloud protocol mitigated the likelihood of hindsight. Fifth, think-aloud protocols are not believed to fully capture nonanalytic (system 1 or pattern recognition) reasoning<sup>14</sup>; therefore, this study may not capture all the types of reasoning tasks residents may use during a patient encounter. Finally, the instructions of the think-aloud protocol may have also limited the ability to capture management tasks if residents worked only to reach a diagnosis.

This study extends previous work to develop a framework for clinical reasoning tasks.<sup>6</sup> Our results suggest that the use of reasoning tasks does not proceed in a sequential or linear manner but functions as more of a nonlinear process. In addition, noteworthy variability occurs among participants in how they use reasoning tasks, including the types of tasks they use and how frequently they use them. This variability among participants provides potential insight into the phenomenon of context specificity—whereby a single physician can see two patients with the same presenting symptoms and findings (who have the same diagnosis) yet develop two different diagnoses because the context has changed. Enhancing the medical community's understanding of clinical reasoning tasks could help mitigate this unwanted physician variance. Further, from our perspective as internists, we have provided suggestions for strengthening the framework to more fully encompass the spectrum of tasks that occur in a clinical encounter.

More research is needed to explore the association between reasoning tasks and diagnostic accuracy. A larger study is necessary to meaningfully determine whether a correlation exists between, on one hand, the types of reasoning tasks used, the frequency of use, and the order of use and, on the other, the impact these have on diagnostic accuracy. Further studies are also necessary to explore the following questions: (1) In what way do resident trainees use the therapeutic

reasoning tasks in practice? (2) How does the use of reasoning tasks change as a resident's expertise increases? (3) How and why do experts use specific reasoning tasks, and how do the tasks differ across contexts? (4) Is the failure to use certain tasks associated with medical error or poorer quality of care? The answers to these and other such questions could provide valuable insight into how to optimize and assess the process of clinical reasoning in future students and physicians. Next steps should include both simulated and in vivo studies with practicing and resident physicians.

*Funding/Support:* This project was supported, in part, by an unrestricted educational grant from MedU/iInTime as well as local intramural grant funding.

*Other disclosures:* None reported.

*Ethical approval:* The authors obtained institutional review board (IRB) approval from the Uniformed Services University of the Health Sciences (protocol number R083VC-03). The Brooke Army Medical Center IRB acknowledged the Uniformed Services University of the Health Sciences IRB protocol and approved the study by memo.

*Disclaimer:* The views expressed in this report are those of the authors and do not necessarily represent the views of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or other federal agencies.

**E. McBee** is assistant professor of medicine, Uniformed Services University of the Health Sciences, based at Naval Medical Center San Diego, San Diego, California.

**T. Ratcliffe** is assistant professor of medicine, University of Texas Health Science Center, San Antonio, Texas.

**M. Goldszmidt** is associate professor of medicine, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada.

**L. Schuwirth** is professor of medicine, Flinders University, Adelaide, Australia.

**K. Picho** is assistant professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**A.R. Artino Jr** is associate professor of preventive medicine and biometrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**J. Masel** is third-year resident, Walter Reed National Military Medical Center, Bethesda, Maryland.

**S.J. Durning** is professor of medicine and pathology, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

## References

- Higgs J, Jones M. Clinical decision making and multiple problem spaces. In: Higgs J, Jones M, Loftus S, Christensen N, eds. *Clinical Reasoning in the Health Professions*. 3rd ed. New York, NY: Elsevier; 2008:3–18.
- Bissessur SW, Geijteman E, Al-Dulaimy M, et al. Therapeutic reasoning: From hiatus to hypothetical model. *J Eval Clin Pract*. 2009;15:985–989.
- Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Health Sci Educ Theory Pract*. 2012;17:65–79.
- Norman G. Research in clinical reasoning: Past history and current trends. *Med Educ*. 2005;39:418–427.
- McGuire CH. Medical problem-solving: A critique of the literature. *J Med Educ*. 1985;60:587–595.
- Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. *Acad Med*. 2013;88:390–397.
- Gibson J. The theory of affordances. In: Shaw R, Bransford J, eds. *Perceiving, Acting and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977.
- Young MF, Barab SA, Garrett S. Agent as detector: An ecological psychology perspective on learning by perceiving-acting systems. In: Jonassen DH, Land SM, eds. *Theoretical Foundations of Learning Environments*. Mahwah, NJ: Erlbaum; 2000:147–173.
- Durning S, Artino AR Jr, Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: Understanding the perspective of the expert's voice. *Med Educ*. 2011;45:927–938.
- Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34:30–37.
- Ericsson KA, Simon HA. Verbal reports as data. *Psychol Rev*. 1980;87:215–251.
- Russo JE, Johnson EJ, Stephens DL. The validity of verbal protocols. *Mem Cognit*. 1989;17:759–769.
- Ericsson K. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Hoffman RR, Feltovich PJ, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York, NY: Cambridge University Press; 2006.
- Ericsson KA, Simon HA. *Protocol Analysis: Verbal Reports as Data*. Rev ed. Cambridge, Mass: MIT Press; 1993.
- Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Acad Med*. 1998;73(10 suppl):S1–S5.
- Durning SJ, Artino AR Jr, Pangaro LN, van der Vleuten C, Schuwirth L. Perspective: Redefining context in the clinical encounter: Implications for research and training in medical education. *Acad Med*. 2010;85:894–901.