Published in final edited form as:

Proc Int Symp Hum Factors Ergon Healthc. 2018 June ; 7(1): 1–7. doi:10.1177/2327857918071001.

PARTICIPATORY BULLSEYE TOOLKIT INTERVIEW: IDENTIFYING PHYSICIANS' RELATIVE PRIORITIZATION OF DECISION FACTORS WHEN ORDERING RADIOLOGIC IMAGING IN A HOSPITAL SETTING

Michael F. Rayo, Chandni Pawar, Elizabeth B.-N. Sanders, Beth W. Liston, and Emily S. Patterson

The Ohio State University Columbus, OH USA

Abstract

Critical Decision Method (CDM), a popular cognitive task analysis (CTA) method, is an in-depth retrospective interview that uses a historical non-routine incident to identify experts' decision-making factors in complex socio-technical settings with high consequences for failure. However, it is challenging to use CDM to make comparisons, including those between experts and trainees. We describe an alternative CTA method used to study physicians' decision making for ordering diagnostic imaging. After being primed with 11 simulated patient scenarios, nine attending and 11 resident physicians were asked to map out and present their decision-making process with a bullseye participatory design toolkit. Interviews were analyzed qualitatively, revealing four common decision factors: diagnostic efficacy, patient safety, organizational constraints, and patient comfort. The bullseye maps were used to quantitatively measure priority differences between these decision factors. Attending and resident physicians both prioritized diagnostic efficacy over the other factors (2.38 vs. 3.71, $p < .01$, and 2.59 vs. 3.52, $p < .01$, respectively), but attending physicians' decisions had a higher proportion of non-diagnostic items (65% vs. 50%, $p = .008$). Our results demonstrate the usefulness of this method in eliciting decision factors for a complex, face-valid task and for identifying differences due to levels of expertise and training.

INTRODUCTION

Cognitive task analysis (CTA) is often used to identify factors considered by experts during decision points in complex, socio-technical settings with high consequences for failure. One CTA method is the Critical Decision Method (CDM), which is an in-depth retrospective interview about a historical non-routine incident that uses a set of cognitive probes in a structured fashion to understand expert judgment, including what factors are detected and considered in selecting actions to make (Klein, Calderwood, & MacGregor, 1989). This method has been proven to be highly useful for many human factors engineers, but there can be some challenges associated with its use. Considerable time may have passed since the incident, each interview is about a different incident, and trainees may not yet have experienced a non-routine event, so it is challenging to compare differences in factors for experts and trainees. We describe an alternative CTA interview method in which a high-fidelity human-in-the-loop simulation is used as a priming mechanism. This is followed by a participatory design session with a bullseye toolkit to facilitate a discussion about strategies

used ‘in situ’. The session is then analyzed qualitatively and quantitatively to understand the relative prioritization of different decision factors. Preliminary observations are conducted to ensure the representativeness of the simulation scenarios and the relevancy of the items in the toolkit.

This method was used to reveal how physicians prioritized decision factors when ordering diagnostic radiologic images in order to inform the design of future evidence-based decision-support technologies. Healthcare providers, including physicians who specialize in providing hospital care, are asked to make increasingly complex decisions that must trade-off short-term and long-term patient goals as well as factors such as cost and reducing lengths of stay (Hackner et al., 2001). However, little is understood about how physicians prioritize these factors. Understanding the inherent tradeoffs between multiple goals when selecting a course of care among many alternatives is critical to decision-making in situations of high complexity and uncertainty (Cook and Woods, 2010). One aspect of these decisions is how to interpret and use evidence-based guidance, especially when the evidence is not strong or it conflicts with a physician’s clinical experience. In order to address areas where evidence is not particularly strong, Fitch and colleagues (2001) developed a standard process to create guidelines on the appropriateness of alternate courses of care. This process has been widely adopted, with Appropriate Use Criteria developed for radiology, cardiology, orthopedics, dermatology, and other specialties (cf., Hendel et al., 2013). The process typically involves convening a panel of clinical experts who distill their experiences and the current evidence into a 9-point appropriateness scale. A procedure’s rating on this scale incorporates both the benefits and risks of performing the procedure. The ratings are intended to guide physicians as they make choices such as what procedures to use to gain insight into patient problems. However, using guidance based upon research evidence is cognitively challenging for physicians. Specifically, Klein et al. (2016) identified cognitive challenges of: 1) characterizing the problems experienced by patients, 2) transparency of strength and recency of underlying evidence, 3) incorporating patient-specific information that differs from characteristics of patients in the research studies (also identified by Holmberg et al., 2015), and 4) taking into account past responses of a patient to evidence-based treatment. Due to these challenges, what a physician ultimately determines to be appropriate will be influenced by many factors for any individual patient. Our past research recently revealed additional difficulties with evidence-based decision-support, finding that decision support systems that have less transparency in how evidence is being used and how competing goals are prioritized are ignored or overridden more often (Rayo et al., 2015). This is especially relevant to diagnostic imaging, in which patient safety has emerged as an important factor due to increased risk of cancer from some imaging modalities due to cumulative exposure over time (Rayo et al., 2014). In addition, there are variations in ordering patterns for physicians based on geographic region (Welch et al., 1993) and experience level, similar to expert-novice differences in other complex settings (Randel, Pugh, and Reed, 1996).

Although participatory toolkits have often been used in participatory design research (Sanders 2002, 2006), and have had recent success in hospital applications including hospital room design (Lavender et al., 2015; Patterson et al., 2017), this approach is novel in that it was used for research about decision-making processes. It facilitated obtaining

quantitative measures that support statistical comparisons across decision factors and cohorts of participants. We propose that this method might be useful as an alternative way of conducting a CTA.

METHODS

Participants

Physicians were recruited from multiple general medicine services in a tertiary care academic hospital. IRB-approved requests for volunteers were sent to all resident and attending hospitalist physicians in the study institution. Inclusion criteria were prior experience prescribing advanced diagnostic imaging and completion of training on using the institution's electronic health record.

Research Design

Data collection.—Participants created and then presented their own personal model of decision-making on the appropriateness of diagnostic ordering. Participants were first asked to use a simulated computer system to review imaging orders of varying appropriateness for 11 patients, determining whether they would keep the order, change it to a different order, or delete the order. In addition to producing its own results, which are reported elsewhere (Rayo et al., 2015), the simulations acted as a priming for realistic ordering behaviors, allowing participants better access to the factors considered in the selection of which imaging procedure to order. A secondary priming was also performed by having the participants create a free list (cf., Flinn 1988) of all of the factors they could think of that were important in ordering diagnostic imaging.

After these two preparatory steps, the participants used a picture and word sticker toolkit to create a visual collage (cf., Sanders et al., 2010). For the collaging session, they kept their free list and were given the stimulus items (i.e., stickers of pictures and words) and the bull's-eye background shown in Figure 1. The stimulus items were chosen based on previous interviews with physicians about how they made imaging decisions. The concentric rings of the bull's-eye had radii of 3.2 cm, 7.1 cm, and 11.0 cm. The grey color of the bullseye was chosen so that it would not visually distract from the stimulus items being placed on it. Directions on the bullseye background instructed participants to create a visual collage describing how they made their imaging decision by placing stimulus items that were more helpful towards the center of the bullseye, and less helpful items towards the outside. Items that were not helpful at all were to be placed on the far right side of the background. Participants could use as many or few of the items as they wished, and could also write items from their list directly on the bullseye background. After creating their visual collage, they were asked to present it to the researcher, who asked follow-up questions as necessary. The entire session was audiotaped and videotaped, and later transcribed by one of the researchers. After each participant session, the distance between the center of the bullseye and the center of each stimulus item was measured.

Data analysis.—Emergent themes were identified in the audio transcriptions, which became the categories of the framework of decision factors. An iterative multiple-pass

abductive analysis was performed: first, each line of the transcripts was thematically coded inductively line-by-line by a single researcher: no *a priori* codes were used in this initial analysis. Statements were then grouped by code, and were each read again by the same researcher to determine underlying themes. Confidence in a given theme increased as more confirmatory statements were found; themes were refined or changed when conflicting or disconfirming statements were found. The resulting themes became the decision factors. Each stimulus item was associated with the decision factor that it was most identified with in the participants' accounts. A second researcher then independently associated the decision factors to the stimulus items. Areas of disagreement between the two researchers were resolved to consensus through discussion with a multidisciplinary team comprised of human factors engineers and clinical experts.

The factors were finally condensed into two categories, separating the primary factor (diagnosis) from all secondary factors. A Pearson Chi-square test of independence was used to assess the association proportion of the type of physician (i.e., attending vs. resident physician) and the usage of diagnostic vs. non-diagnostic stimulus items. A two-way ANOVA and Games-Howell test of simultaneous differences were used to calculate differences in stimulus item radius due to the type of physician and whether the item was in the diagnostic or non-diagnostic category. Bonferroni corrections for multiple simultaneous comparisons were used to set the threshold for significance where appropriate.

Data visualizations were created for each of the four decision categories. Stimulus items on each physician's visual map were transformed into semi-transparent ovals color-coded to each of the four decision categories. These ovals were superimposed on each other so that each resultant visualization showed all physicians' responses for each category. The visualizations were then visually examined and compared by the research team.

This study was approved by The Ohio State University Institutional Review Board.

RESULTS

Participants

Eleven of the physicians were resident physicians, and the remaining 9 were attending physicians. Participants' years of experience since graduating from medical school ranged from 1 to 17 years. The average years of experience for attending and resident physicians were 10.8 and 2, respectively. The age of the participants ranged from 23 to 43 years. The average ages for attending and resident physicians were 36.6 and 27.9, respectively. There were 11 males and 9 females.

Grounded Theory and Visual Analysis Themes

The Grounded Theory and visual analyses together revealed four themes: diagnostic efficacy, patient safety, organizational constraints, and patient comfort. The visualization of stimulus items across all four themes is found in Figure 2. For all participants, their primary factor was within the theme of diagnostic efficacy. These items are tightly packed in the center circle, designating that they are the most helpful. Physicians referred to these topics as "the most important thing" or "the first thing I think of".

For a test to have high diagnostic efficacy, four conditions must be met. First, the test will likely provide information that will improve the patient's ultimate outcome. This is highly contextual to the physician's current impression of the patient's condition and the likely diagnoses. It also depends on the physician's confidence in their current diagnosis: a high degree of confidence in diagnosis will make further testing less desirable. Finally, the test results must be amenable to further action:

The best decision ... would be looking at what are the symptoms, what are you looking for in the patient, and then look at the natural history of what you're working up, and tell you what the best test is.

(P2, attending physician, 39 years old, 15 years of experience)

If my diagnosis were borderline, then an image would help with that.

(P6, resident physician, 31 years old, 3 years of experience)

Second, the time to administer the test is low enough so that the wait time to receive and interpret test results does not adversely affect the patient's health. If the patient's condition does not require immediate action, the physician has more time to evaluate care plan options.

How sick is the patient, how much time do I have to think about it.

(P3, resident physician, 30 years old, 4 years of experience)

The type of test, if it's going to help you make your diagnosis quicker and the patient is sick and needs that diagnosis quick, because it will help the patient's management. You see in the ED that they give STAT head CT's all the time, because it will give them a quick yes/no.

(P16, resident physician, 25 years old, 1 year of experience)

Third, results of the test can be trusted. The test has sufficient historical sensitivity and specificity of confirming or disconfirming the physician's likely diagnoses. Again, this is highly contextual to what the physician is looking for in the test.

I always need to know how likely the test is to actually detect that [that I'm looking for].

(P7, attending physician, 34 years old, 8 years of experience)

What is most likely diagnosis, what am I looking for - is it soft tissue, or vascular, or bone ... That's my highest priority.

(P17, resident physician, 26 years old, 1 year of experience)

Fourth, the test must have the potential to change the prescribed course of care. In some situations, the care plan for a positive test result is the same for a negative test result. In other situations, if something is found in the image, how likely can the team fix it?

If I find something, is there something I can do about it? Meaning, if there's nothing I can do about it, do I really care that I found it...

(P7, attending physician, 34 years old, 8 years of experience)

As we go forward, is this test going to change a treatment? One of the reasons why in the sinusitis case is not that exciting is that if you get a scan and there's congestion of sinuses you'll probably give antibiotics. If their sinuses aren't that impressive, you're probably still going to give them a course of antibiotics.

(P2, attending physician, 39 years old, 15 years of experience)

Secondary factors for this decision were within the themes of patient safety, organizational constraints, and patient comfort. The layout for these secondary factors was markedly different from diagnostic efficacy, which can be seen in Figure 2. Each was more diffuse towards the outer rings. Participants generally spoke of them as important but not as important as diagnostic efficacy.

Less helpful but still helpful are things like radiation - if there are two and I could do either one, I tend to go with the one that is less. Same with cost. They're considerations but not major considerations because they're not going to have acute effects - they shouldn't be the driving force behind your decision in the short-term.

(P20, resident physician, 30 years old, 2 years of experience)

Patient safety involved weighing the potential unintended side effects of the imaging with the benefits of the information the test provides. The two most noted considerations were the short-term risk of kidney disease for tests with contrast, and the long-term risk of radiation exposure for tests with ionizing radiation. The layout of the patient safety items was concentrated in the middle ring, designating moderate importance, although many physicians placed safety items in the center ring as well. Often, they were introduced as the "next thing I think of" or "almost equally important."

Equally important is the safety for that test, so I tie safety with considering the radiation and other risks of the test like contrast-induced nephropathy with the patient characteristics for safety, like comorbidities, more at risk for contrast, or if you're pregnant, you have an increased sensitivity to radiation, so those are all linked to me.

(P7, attending physician, 34 years old, 9 years of experience)

For some, however, these aspects of safety played no role in their decision.

The sad thing is that I don't always think about radiation dose.

(P13, attending physician, 31 years old, 6 years of experience)

I never think about radiation dose. I would have no idea.

(P14, resident physician, 30 years old, 1 year of experience)

Organizational constraints, sometimes called blunt end constraints (Hollnagel, Woods, and Leveson, 2007), are the collection of pressures exerted on the physicians from the rest of the organization. Among them are cost, legal liability, clinical policies, or past funding decisions that now affect the timely availability of specific test. These items were also concentrated in the middle ring, but had fewer placements in the center ring. They were often introduced as things that “you also have to think about”, or that “I don’t want to think about, but I do.”

Cost is certainly a factor, but it's a little bit less of a factor as long as the test is clinically appropriate.

(P12, attending physician, 41 years old, 16 years of experience)

Backlog refers to MRI, it's hard to get them done, it takes time, so rapid is important, the ability to get an appropriate test done.

(P4, attending physician, 38 years old, 10 years of experience)

Patient comfort is a factor contextualized by the likelihood that the patient will be able to tolerate the recommended imaging. Considerations for patient comfort for this decision include the patient’s preferences for imaging, difficulties lying still without sedation, and claustrophobia. These items were not always mentioned by the physicians, and were found in almost equal measure in the outer two rings.

I put patient preference on the borderline because sometimes a patient says “I have too much anxiety, I can't get an MRI,” I'm able to bargain with them to tolerate the procedure if I feel strongly about the test.

(P11, resident physician, 25 years old, 1 year of experience)

Although this categorization of primary and secondary factors implies a linear and strictly hierarchical decision-making model, multiple physicians commented that patient context can make a big difference in the relative priorities of the different themes.

Put radiation in the middle ... You want to avoid irradiating someone as much as you can ... You have to think about it, but if the person needs a test, they need a test.

(P5, resident physician, 29 years old, 2 years of experience)

Time is only important when it's important.

(P6, resident physician, 31 years old, 3 years of experience)

Statistical Analysis

The four decision factors resulting from the grounded theory and visual analysis were grouped into two larger categories (diagnostic and non-diagnostic) in order to investigate the relationship of experience and decision-making preferences for primary and secondary factors. Table 1 compares the average number of items placed per physician, the proportion of diagnostic items, and the distance from center (i.e., radius) of diagnostic and non-diagnostic items for attending and resident physicians. Attending physicians' use of diagnostic factors was proportionally less than resident physicians (45/130 vs. 89/179, $p = .008$). Diagnostic factors were placed closer to the center of the bullseye than non-diagnostic factors for both attending (2.38 vs. 3.71, $p < .01$) and resident physicians (2.59 vs. 3.52, $p < .01$).

DISCUSSION

Using the participatory bullseye toolkit, physician participants across all levels of experience consistently communicated that the diagnostic efficacy of a test for a given patient's situation was more important than other factors in making decisions about ordering diagnostic imaging. They also consistently described diagnostic efficacy in terms of four subcomponents: that the test provides *relevant*, *timely*, and *trusted* answers that have the ability to *impact the care plan*. Three other important factors were also identified for all levels of expertise: patient safety, organizational constraints, and patient comfort. This consistency of definition and prioritization is perhaps not surprising, as clinicians have repeatedly shown that they will adopt shared prioritization hierarchies, even when there is no coordinated effort or training to inculcate that hierarchy. In one study, 87% of nurses described that during observed decisions between two options on acute care units, they chose to address imminent clinical concerns rather than other options such as documenting the effectiveness of medications to reduce pain (Patterson, Ebright, Saleem, 2011). In another observational study, clinicians ignored or delayed addressing clinical reminders in the electronic health record during an outpatient visit with a patient, explicitly deprioritizing this work below that of what they perceived to be their core work (Saleem et al., 2005).

One interesting difference between attending physicians and resident physicians was how often the non-diagnostic factors were mentioned. Resident physicians discussed non-diagnostic topics less often than attending physicians, perhaps indicating a less evolved understanding of the factors and relationships between them.

In addition to the insights that were derived about this decision, this study demonstrated that participatory methods were a pragmatic way to better understand work in the Medical Center. This has opened the door to subsequent projects. Two recent examples are (1) using a three-dimensional toolkit to understand how physicians and nurses anticipate and recognize patient decompensation and (2) use an interactive toolkit to design new alarms to redirect clinicians' attention to more urgent medical situations.

Our study had a number of limitations. In order to enable the rich discussion and to take the time to collect the detailed information, it would be impractical to have a large number of participants. Therefore, only 20 study participants were recruited from a single institution in a single specialty (hospital medicine). As a result, findings may not generalize to other institutions or other medical specialties. In order to develop a toolkit that included probes that covered the likely space of possible factors for consideration during imaging ordering, preliminary observations and interviews were conducted. For the statistical analysis, the residuals for the ANOVA were slightly non-normal in distribution, but sufficient data were analyzed to warrant the claim of statistical significance without needing to do a logarithmic transformation of the data.

CONCLUSIONS

In summary, we believe that this method provided useful information about the relative prioritization of factors considered with differential importance by physicians with varying levels of training. Requiring physical placement of each item to denote its relative importance was easily understood and forced measurable differences in a standardized way. Analysis of the description of the reason for the placement of items yielded a rich source of how individual patient considerations could potentially change the relative importance of the factors. Therefore, the interviews afforded a rich source of insights about the nature of expertise, including how patient context might be used to appropriately tailor evidence-based guidance. Similar to other CTA methods, our insights could inform the design of clinical decision support tools as well as training content for medical students and resident physicians.

ACKNOWLEDGMENTS

We thank Ted Allen, PhD, for statistical support during data analysis. We also thank the physicians at the Ohio State Medical Center for being willing participants and helping us recruit their colleagues for this study. This research was supported by the Institute for the Design of Environments Aligned for Patient Safety (IDEA4PS) at The Ohio State University which is sponsored by the Agency for Healthcare Research & Quality (P30HS024379). The authors' views do not necessarily represent the views of AHRQ.

REFERENCES

- Cook RI, & Woods DD. (2010). Goal conflicts In Behind Human Error. Ashgate Publishing Company , 123–142 .
- Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR. The RAND/UCLA Appropriateness Method User's Manual. 2001
- Flinn J, (1998). Freelists, ratings, averages, and frequencies: Why so few students study anthropology Using Methods in the Field: A Practical Introduction and Casebook. de Munck VC, ; and Sobo EJ, eds, 85–96 .
- Hackner D, Tu G, Braunstein GD, Ault M, Weingarten S, Mohsenifar Z. The value of a hospitalist service: Efficient care for the aging population? CHEST Journal. 2001; 119(2):580–589.
- Hendel RC, Patel MR, Allen JM, Min JK, Shaw LJ, Wolk MJ, Doherty JU. Appropriate use of cardiovascular technology: 2013 ACCF appropriate use criteria methodology update: A report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force Journal of the American College of Cardiology. 2013; 61(12):1305–1317. [PubMed: 23433633]
- Hollnagel E, Woods DD, , & Leveson N. ; (2007). Resilience Engineering: Concepts and Precepts. Ashgate Publishing, Ltd ..

- Holmberg C, Waters EA, Whitehouse K, Daly M, McCaskill-Stevens W. My lived experiences are more important than your probabilities: The role of individualized risk estimates for decision making about participation in the study of Tamoxifen and Raloxifene (STAR) Medical Decision Making. 2015; 35(8):1010–1022. . DOI: 10.1177/0272989X15594382 [PubMed: 26183166]
- Klein GA, Calderwood R, Macgregor D. Critical decision method for eliciting knowledge IEEE Transactions on Systems, Man, and Cybernetics. 1989; 19(3):462–472.
- Klein DE, Woods DD, Klein G, Perry SJ. Can we trust best practices? Six Cognitive challenges of evidence-based approaches Journal of Cognitive Engineering and Decision Making. 2016 1555343416637520.
- Lavender SA, Sommerich CM, Sanders EB-N, Evans KD, Li J, Umar RZR. , et al. (2015). Med/Surg Patient Room Layouts Identified Through a Participatory Design Process with Hospital Staff Presented at the 19th Triennial Congress of the International Ergonomics Association , Melbourne .
- Patterson ES, Sanders EB-N, Sommerich CM, Lavender SA, Li J, Evans KD. Meeting Patient Expectations During Hospitalization: A Grounded Theoretical Analysis of Patient-Centered Room Elements: HERD: Health Environments Research & Design Journal. 2017; doi: 10.1177/1937586717696700
- Patterson ES, Ebright PR, Saleem JJ. Investigating stacking: How do registered nurses prioritize their activities in real-time? International Journal of Industrial Ergonomics. 2011; 41(4):389–393.
- Randel JM, Pugh HL, Reed SK. Differences in expert and novice situation awareness in naturalistic decision making International Journal of Human-Computer Studies. 1996; 45(5):579–597.
- Rayo MF, Kowalczyk N, Liston BW, Sanders EB-N, White S, Patterson ES. Comparing the effectiveness of alerts and dynamically annotated visualizations (davs) in improving clinical decision making Human Factors: The Journal of the Human Factors and Ergonomics Society. 2015; 57(6):1002–1014. . DOI: 10.1177/0018720815585666
- Rayo MF, Patterson ES, Liston BW, White S, Kowalczyk N. Determining the rate of change in exposure to ionizing radiation from CT scans: A database analysis from one hospital Journal of the American College of Radiology. 2014; (7):11.
- Saleem JJ, Patterson ES, Militello L, Render ML, Orshansky G, Asch SM. Exploring barriers and facilitators to the use of computerized clinical reminders Journal of the American Medical Informatics Association. 2005; 12(4):438–447. [PubMed: 15802482]
- Sanders EB-N. From user-centered to participatory design approaches Design and the Social Sciences: Making Connections. 2002:1–8.
- Sanders EB-N. Design research in 2006 Design Research Quarterly. 2006; 1(1):1–8.
- Sanders EB-N, Brandt E, Binder T. A framework for organizing the tools and techniques of participatory design In: ACM Press. New York, New York, USA: ACM Press; 2010 :195–198 . doi: 10.1145/1900441.1900476.
- Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States New England Journal of Medicine. 1993; 328(9):621–627. [PubMed: 8429854]

Figure 1:
Participatory toolkit words (A), pictures (B) and bull's-eye background (C)

Figure 2:

Aggregate visualizations depicting the locations of all stimulus items used for each decision factor

Table 1:

Comparison between attending and resident physicians of average radius and number of diagnostic and non-diagnostic topics

	Attending	Resident	p
Avg. items used per physician	14.4	16.2	
Proportion diagnostic items	35%	50%	.008
Avg. radius diagnostic	2.38	2.59	
Avg. radius non-diagnostic	3.71	3.52	
p-value radius diag. vs. non-diag.	<.01	<.01	