

# Assessing Use of Cognitive Heuristic Representativeness in Clinical Reasoning

Velma L. Payne, MS, MBA, MS<sup>1</sup>, Rebecca S. Crowley, MD, MS<sup>1, 2, 3</sup>

<sup>1</sup>Department of Biomedical Informatics, University of Pittsburgh School of Medicine and

<sup>2</sup>Department of Pathology, University of Pittsburgh School of Medicine and

<sup>3</sup>Intelligent Systems Program, University of Pittsburgh School of Arts and Sciences

## Abstract

*We performed a pilot study to investigate use of the cognitive heuristic Representativeness in clinical reasoning. We tested a set of tasks and assessments to determine whether subjects used the heuristics in reasoning, to obtain initial frequencies of heuristic use and related cognitive errors, and to collect cognitive process data using think-aloud techniques. The study investigates two aspects of the Representativeness heuristic - judging by perceived frequency and representativeness as causal beliefs. Results show that subjects apply both aspects of the heuristic during reasoning, and make errors related to misapplication of these heuristics. Subjects in this study rarely used base rates, showed significant variability in their recall of base rates, demonstrated limited ability to use provided base rates, and favored causal data in diagnosis. We conclude that the tasks and assessments we have developed provide a suitable test-bed to study the cognitive processes underlying heuristic errors.*

## Introduction

Medical errors are among the top ten causes of death in the United States<sup>1</sup>. Motivated by the report *To Err is Human* the healthcare industry is addressing preventable medical errors.<sup>1</sup> Risser defines *mistakes* as incorrect actions caused by misclassifying a situation or failing to take into account relevant decision factors.<sup>2</sup> Studies have shown that cognitive errors underly most diagnostic errors made in clinical decisions in the emergency room.<sup>3-5</sup> By understanding the processes that result in cognitive errors, researchers may be able to better design systems and methods to limit these errors.

Human information processing and prospect theories suggest that people have limited cognitive capacity.<sup>6</sup> Consequently, people rely on cognitive heuristics to reduce complex input data to manageable dimensions. The management of data is especially important within the medical environment. Clinicians process a great deal of information often under conditions of uncertainty and stress.

We performed a pilot study to test a methodology for studying the cognitive heuristic *Representativeness* as it is applied in clinical reasoning. We developed a

set of tasks and instruments to (1) measure the frequency of heuristic use, and (2) obtain cognitive process information as subjects are using the heuristic.

## Background

The Representativeness heuristic was first discussed by Kahneman and Tversky in the early 1970's.<sup>7</sup> Matching data to mental models stored in memory is a common way of thinking about Representativeness. Tversky and Kahneman identified four basic situations in which Representativeness is invoked which we describe below.<sup>8</sup>

### Representativeness Type 1 – Judging by Perceived Frequency

During clinical reasoning, Representativeness type 1 (R1) is invoked by assessing a patient's problem in terms of a relative frequency distribution of a specific variable in a specific population. For example, using R1, a young child with loss of motor skills and listlessness would more likely be diagnosed with Tay-Sachs disease if the parents are both Ashkenazic Jews (specific variable), because of the perceived frequency of the Tay-Sachs disease in the Ashkenazic Jewish population (specific population).

### Representativeness Type 2 – Representativeness as Similarity to a Prototype

Representativeness type 2 (R2) is based on similarity to a prototype. During the diagnostic process a patient's features are matched to mental representations. For example, using R2, a patient who has symmetrical, severe metacarpophylangeal redness and tenderness that worsens in the evening and a positive Rheumatoid Arthritis factor (RF) will be diagnosed with Rheumatoid Arthritis, because they are similar to a prototype (mental representation) of patients with this disease.

### Representativeness Type 3 – Representativeness as Variability

Representativeness type 3 (R3) is similar to type 2, but is used to assess a set of patients against the entire population. Judgment is based on the similarity of a subset of the population to the population as a whole. When using this type, one

must consider the fact that small samples are not adequate indicators of health problems in the larger population and not place inappropriate confidence in small samples even if they resemble a larger population. For example, when reasoning about the possibility of a bioterrorism event, a clinician would use R3 when assessing whether the variability in serum potassium among a cohort of ill patients (subset) is representative of patients who unknowingly ingest potassium hydroxide following poisoning of a water supply (entire population).

#### **Representativeness Type 4 – Representativeness as Causal Beliefs**

When using Representativeness type 4 (R4) during the diagnostic process, a patient's problem is perceived in terms of a causal system. Judgment is made by theorizing about cause and effect relationships; i.e. event x is the cause of disease y. For example, in a patient with hematuria and a history of benzidine exposure, the diagnostic consideration of bladder carcinoma may arise quickly due the causal relationship of the toxin to the neoplasm.

#### **Heuristics and Cognitive Errors**

The application of any cognitive heuristic, such as Representativeness, is a double-edged sword. In many cases, use of the heuristic enables accurate and rapid pattern matching in support of classification problem solving. But in some cases, the application of these heuristics leads to serious errors.

Following the example described for Representativeness Type 2, an elderly patient with joint pain and a positive RF factor might be diagnosed with Rheumatoid Arthritis by a novice clinician, because the diagnostician is unaware that RF is positive in a subset of patients with no disease, which prevents the inference that the evidence is more consistent with Osteoarthritis.

Using the example described for Representativeness Type 4, a clinician may overestimate the likelihood of bladder cancer in a patient with symptoms and signs that could be either prostate cancer or bladder cancer because the patient has a history of working in the oil industry. In this case, the physician may ignore base rate data that prostate cancer is more common than bladder cancer and rely more heavily on the causal data of carcinogen exposure.

Despite the common use of heuristics, the ability of expert diagnosticians to produce accurate classification is rarely limited by heuristic errors. We

hypothesized that tasks that emphasize the acquisition of new diagnostic skills provided a good test bed for studying use and misuse of heuristics.

#### **Research Objectives and Questions**

The long term objective of this work is to examine the cognitive processes underlying heuristic errors and to develop techniques to reduce these errors. This pilot study was designed to test our methodology for studying these cognitive processes and to demonstrate that these tasks result in R1 and R4 errors. We selected these heuristics for study because they are known to be associated with errors and seem most immediately amenable to correction using cognitive forcing strategies. Thus, an important overall objective of this work is to determine that we can reproduce these previously described cognitive phenomena using tasks in a clinical domain, and use cognitive methods and assessments to measure them. Analysis of cognitive data will only be valid if we can first demonstrate that we are indeed studying clinicians exhibiting errors produced by application of these heuristics. Towards this objective, we asked:

**Question 1:** Do clinicians make errors in judging by perceived frequency? (Errors secondary to R1)

**Question 2:** If base rates and causal information are both available, do clinicians make errors in weighing causal data more heavily? (Errors secondary to R4)

Another objective of this research is to develop a coherent methodology for studying the cognitive dimension of errors produced by application of these heuristics. To study the cognitive processes underlying heuristic errors, as well as variability among individuals, it is necessary to dissect the phenomenon into more discrete cognitive processes for further study. In support of the second objective we asked the following research questions:

**Question 3:** Do clinicians consciously infer base rates when learning new schemata from a set of training cases that simulate a patient population? If so, do they use these base rates in clinical reasoning?

**Question 4:** If they do not consciously infer base rates when learning new schemata, can they retrospectively provide accurate base rates from the training cases when prompted to do so?

**Question 5:** If clinicians do not use base rates over causal data naturally, do they consider base rates in reasoning if the rates are made available to them?

**Question 6:** Do individuals show differences in frequency of errors related to heuristic use as a function of their domain of expertise?

## Methods

### Subjects

Physician residents including four pathology residents and four internal medicine residents of varying post-graduate year were subjects. Subjects were recruited using email. Each subject performed all tasks in a single session (about 2 hours). All subjects received payment for participation. Use of human subjects was conducted under an exempt University of Pittsburgh IRB protocol (#PRO07080170).

### Study Tasks, Instruments and Procedures

#### Task 1: Clinical Vignette Assessments

To assess research questions 1, 2, 5 and 6, subjects completed an assessment consisting of clinical vignettes followed by a set of possible answers (Figure 1). Responses were worded such that they included a correct answer, an incorrect answer, and one or more answers that would be consistent with a specific cognitive bias (either R1 or R4). Two forms were used. The first form included topics in general medicine and was completed by all subjects. This instrument was developed by other investigators and validated in a previous study.<sup>10</sup> The second form contained questions patterned after the first form but using the domain of general pathology. Because it uses a specialty domain, the second form was completed only by the pathology residents. Subjects did not think-aloud when completing the vignettes.

The vast majority of non-trauma patients with respiratory failure who were treated in the emergency room had a history of chronic obstructive lung disease (COPD). Many of these patients had end-stage COPD and were already receiving continuous home oxygen prior to the episode of respiratory failure. The average age of patients with COPD is 67. Which of the following is more probable for a patient admitted from the emergency room with respiratory failure that is not due to trauma?

- The patient has COPD and is on home O<sub>2</sub>
- The patient has COPD
- The patient has COPD and is over 67
- The patient is over 67 and is on home O<sub>2</sub>

Figure 1. Sample Clinical Vignette

#### Task 2: Feature-Diagnosis Card Simulation with Think-Aloud Protocols

To address research question 1 and 3, feature-diagnosis cards of skin diseases were developed to create a simulated population of patients in order to determine if base rates are inferred when evaluating a population, and to ascertain if inferred rates impact

clinical reasoning and diagnosis. Three training sets of 20 cases containing both features and diagnoses alternated with three test sets of eight cases that contained only features, for a total of 84 cards. Each card represented a different patient and contained patient identification, a set of features and a diagnosis (Figure 2). Cards contained repeated diseases with a frequency representative of disease base rates reported in the literature, i.e. out of the total number of cards, the percentage of cards for Disease X was equal to the incidence of the disease derived from the literature. Prior to performing tasks, subjects were trained on techniques of thinking aloud<sup>9</sup>. Subjects were asked to think aloud as they assessed training set cards and made a diagnosis for each card in the test sets. Verbal protocols were captured on audiotape, and were transcribed and coded using standard techniques<sup>9</sup> to determine if subjects inferred and used base rates during diagnosis of test cases.

**Patient 1 – A.S. presents with**

- Eosinophils in blister
- Eosinophilic dermal inflammatory infiltrate
- Sub-epidermal blister

**Diagnosis: Bullous Pemphigoid**

Figure 2. Sample Feature-Diagnosis Card

#### Task 3: Recall of Base Rates

To address research question 4, after evaluating all feature-diagnosis cards, subjects were asked to provide base rates for each disease they encountered in the patient population represented by the cards. We calculated the statistical correlation to determine how closely rates provided by subjects correspond to rates inherent in the card set.

#### Task 4: Disease Pattern and Causality Cards with Think-Aloud Protocols

To address research question 2 and 5, using the same method described for Task 2, all subjects reviewed a second card set that identified disease patterns and base rates; as well as causal data cards that provided information on causes and common patient population for diseases; and then assessed clinical scenarios that referenced base rates, causal data, or both. Disease pattern cards (Figure 3) contained Venn diagrams reflecting common disease features, features unique to diseases, and base rates of the disease.

Disease causal cards (Figures 4) provided information designating the cause of skin diseases (if known), as well as patient populations affected by the disease. Think-aloud verbal protocols were captured, transcribed and coded to assess the extent clinicians

used base rates and causal data when this information was known to them.

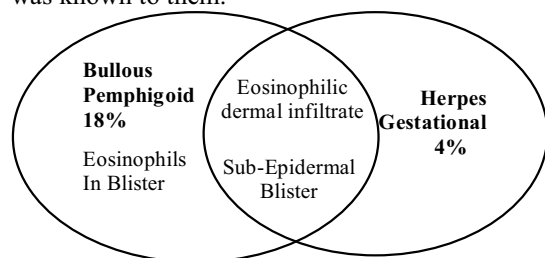


Figure 3. Sample Disease Pattern Card

Bullous Pemphigoid is a disorder in which antibodies produced by the immune system attack a thin layer of connective tissue in the skin (basement membrane). It is not certain what causes antibodies to attack healthy tissue in the body. If the delicate basement membrane becomes inflamed, fluid blisters can form. Bullous Pemphigoid usually occurs in elderly persons and is rare in young people.

Figure 4. Sample Disease Causal Card

## Results

### Question 1 – Errors due to use of R1 heuristic

Based on results of Task 1 (R1 vignettes), subjects selected the correct answer in 59% of vignettes, and selected the incorrect answer in 41% of vignettes. Of the 13 total incorrect answers, subjects selected the representativeness answer 76% of the time. The data suggest that errors due to R1 are a result of base rate neglect.

### Question 2 - Errors due to use of R4 heuristic

Results of Task 1 (R4 vignettes) show subjects selected the correct answer in 84% of vignettes, and selected the incorrect answer in 16% of vignettes. Of the 5 incorrect answers, subjects selected the answer weighted towards causal data 100% of the time. The data suggest that the heuristic bias resulting from base rate neglect accounts for diagnostic errors.

### Question 3 - Inference of Disease Base Rates

Results from Task 2 (feature-diagnostic card simulation) show that when assessing a population of simulated patients, subjects did not infer disease base rates. The mean protocol statements related to disease frequency ( $3 \pm 3$ ) in proportion to mean total number of statements ( $456 \pm 120$ ) accounts for fewer than 2% of think-aloud statements. Occasional statements related to frequency were observed, such as 'I've seen this disease quite often in these patients' or 'This is the first time I have seen this disease'; but subjects never explicitly used frequency of diseases in the training set to help them determine a diagnosis in the test set.

### Question 4 - Retrospective Recall of Base Rates

The overall correlation of estimated disease base rates provided by the subjects to actual base rates inherent in the card sequence (Task 3) ranged from -.03 to +0.85, with a mean of +0.35, suggesting that on average, subjects were relatively inaccurate in remembering the distribution of the diseases within the population. However, there was substantial variation in subjects' abilities to recall or construct base rate information.

### Question 5 - Use of Base Rate and Causal Information

Based on the results of Task 4 (disease pattern and causality cards), when subjects were provided with disease base rate and causal information, they used causal information over base rates 82% of the time.

### Question 6 - Differences Based on Domain

For subjects who completed the general medical and pathology specific vignette assessments, we observed that errors related to R1 and R4 were also evident in domain specific vignettes. For the domain specific R1 vignettes (Task 1), pathology residents selected the correct answer in 63% of the vignettes, and selected the incorrect answer in 38% of the vignettes. Of the 6 total incorrect answers, subjects selected the representativeness answer 75% of the time.

Of the domain specific R4 vignettes (Task 1) subjects selected the correct answer in 75% of the vignettes, and selected the incorrect answers in 25% of the vignettes. Of the 4 total incorrect answers, subjects selected the answer weighted towards causal data in two instances (50%) and the answer using both causal and base rate information in two instances (50%).

## Discussion

The purpose of this study is to (1) pilot test a methodology for studying the cognitive processes associated with heuristic errors, (2) validate that performance of the tasks we assigned is associated with use of R1 and R4 heuristics in the target subject population, and (3) determine to what extent subjects make cognitive errors associated with application of these heuristics.

Using cognitive methods and assessment techniques to measure the use of cognitive heuristics in a clinical domain, this study revealed clinicians frequently use Representativeness types 1 and 4 inappropriately during clinical diagnostic reasoning. Inappropriate use of R1 and R4 as a result of the base rate neglect

phenomenon in which base rates were ignored led to medical errors. This finding was confirmed by the lack of clinicians' inference of base rates while assessing patients, the inability to provide accurate base rates immediately after assessing an aggregate of patients, and the use of causal data over provided base rates. Our study is commensurate with other studies in showing clinicians' use cognitive heuristics when assessing clinical data.<sup>9-11</sup> We conclude that resident physicians reasoning in general medicine and pathology, using tasks performed in this study, commonly use cognitive heuristics inappropriately, with characteristics similar to those previously established in the literature within medical and non-medical domains.

### Limitations

Even though we were able to demonstrate that clinicians exhibit errors when utilizing cognitive heuristics in clinical situations, a limitation of our study was the use of a small number of subjects and the descriptive nature of a pilot study. We are unable to generalize our results to the general population of clinicians. Nevertheless, the frequency of these cognitive biases in the task environment for this small sample leads us to believe that a larger sample could provide useful information on the frequency of Representativeness errors in medical reasoning.

### Future Work

Future research will focus on analyzing the think-aloud protocols generated during tasks 2 and 4, and in comparing cognitive processes observed in these tasks to recall data in task 3 and data from standardized instruments in task 4. Data from think-alouds provide a method for inferring the mental models that clinicians construct as they learn new schemata (Task 1), and as they interpret frequency data (Task 4). Why are some clinicians better able to recall and use base rate data than others? Do characteristics of mental models constructed for representing evidence-feature relationships result in predictable limitations susceptible to heuristic errors? Analysis of think-aloud protocols will be correlated with data described in this manuscript to answer these questions. In the future, we anticipate using these insights to guide development of computer based educational systems that debias clinicians from inappropriate use of cognitive heuristics such as Representativeness.

### Conclusion

We have demonstrated a method for capturing cognitive data related to heuristic use in clinical reasoning. This work is a first step towards

understanding errors that result from heuristic use through cognitive task analysis. Future research will focus on correlating the findings with cognitive processes, with the aim of developing specific remediation strategies for reducing the frequency of these common cognitive errors.

### Acknowledgements

Our thanks go to Catherine G. Ferrario, PhD, University Of St. Francis, for permission to use the clinical vignettes. This work was supported by the National Library of Medicine University of Pittsburgh Biomedical Informatics Training Program Grant (T15 LM007059-2).

### References

1. Institute of Medicine, Washington D.C. To Err is Human. Building A Safer Health System. National Academy Press. 2000.
2. Risser T, Rice M, Salisbury M, Simon R, Jay GD, Berns SD. The Potential for Improved Teamwork to Reduce Medical Errors in the Emergency Department. *Annals of Emergency Medicine*. 1999;34(3):373-83.
3. Croskerry P. The Cognitive Imperative: Thinking About How We Think. *Academic Emergency Medicine*. 2000 November;7(11):1223--31.
4. Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Academic Emergency Medicine*. 2002;9:1184-12-4.
5. Croskerry P. Cognitive Forcing Strategies in Clinical Decision Making. *Academic Emergency Medicine*. 2003;41:110-20.
6. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*. 1979;XLVII:263-91.
7. Kahneman D, Tversky A. Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*. 1972;3:430-54.
8. Tversky A, Kahneman D. Causal Schemata in Judgments Under Uncertainty. In M Fishbein (Ed), *Progress in Social Psychology*. 1980; Hillsdale, NJ: Erlbaum.
9. Ericsson KA, Simon HA. *Protocol Analysis (Revised Edition); Overview of Methodology of Protocol Analysis*. 1993.
10. Ferrario CG. *The Association of Clinical Experience and Emergency Nurses' Diagnostic Reasoning*. PhD Dissertation, Ann Arbor, MI: Rush University; 2001.
11. Casscells W, Schoenberger A, Graboys T. Interpretation of physicians of laboratory results. *New England Journal of Medicine*. 1978;299(18):999-1000.
12. Christensen-Szalanski J, Bushyhead J. Physician's use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance*. 1981;7(4):928-35.