

# BRIEF REPORT: Beyond Clinical Experience: Features of Data Collection and Interpretation That Contribute to Diagnostic Accuracy

Mathieu R. Nendaz, MD, MHPE,<sup>1,2</sup> Anne M. Gut, PhD,<sup>1</sup> Arnaud Perrier, MD,<sup>2</sup> Martine Louis-Simonet, MD,<sup>2</sup> Katherine Blondon-Choa, MD,<sup>2</sup> François R. Herrmann, MD, MPH,<sup>3</sup> Alain F. Junod, MD,<sup>2</sup> Nu V. Vu, PhD<sup>1</sup>

<sup>1</sup>Unit of Development and Research in Medical Education, University of Geneva, Faculty of Medicine and Geneva University Hospitals, Geneva, Switzerland; <sup>2</sup>Department of Internal Medicine, University of Geneva, Faculty of Medicine and Geneva University Hospitals, Geneva, Switzerland; <sup>3</sup>Department of Rehabilitation and Geriatrics, University of Geneva, Faculty of Medicine and Geneva University Hospitals, Geneva, Switzerland.

**BACKGROUND:** Clinical experience, features of data collection process, or both, affect diagnostic accuracy, but their respective role is unclear.

**OBJECTIVE, DESIGN:** Prospective, observational study, to determine the respective contribution of clinical experience and data collection features to diagnostic accuracy.

**METHODS:** Six Internists, 6 second year internal medicine residents, and 6 senior medical students worked up the same 7 cases with a standardized patient. Each encounter was audiotaped and immediately assessed by the subjects who indicated the reasons underlying their data collection. We analyzed the encounters according to diagnostic accuracy, information collected, organ systems explored, diagnoses evaluated, and final decisions made, and we determined predictors of diagnostic accuracy by logistic regression models.

**RESULTS:** Several features significantly predicted diagnostic accuracy after correction for clinical experience: early exploration of correct diagnosis (odds ratio [OR] 24.35) or of relevant diagnostic hypotheses (OR 2.22) to frame clinical data collection, larger number of diagnostic hypotheses evaluated (OR 1.08), and collection of relevant clinical data (OR 1.19).

**CONCLUSION:** Some features of data collection and interpretation are related to diagnostic accuracy beyond clinical experience and should be explicitly included in clinical training and modeled by clinical teachers. Thoroughness in data collection should not be considered a privileged way to diagnostic success.

**KEY WORDS:** clinical reasoning; clinical data collection; experience; expertise; medical education; internal medicine.

DOI: 10.1111/j.1525-1497.2006.00587.x

J GEN INTERN MED 2006; 21:1302-1305.

Studies in cognitive psychology have described the processes of clinical reasoning, the organization of memory, and the mental representations of knowledge.<sup>1,2</sup> Characteristics influencing data collection or recognition have been well documented in visual clinical disciplines like dermatology, or in cases for which the patient's physical appearance leads to the diagnosis.<sup>3-6</sup> For situations containing less visible data, previous studies including experienced physicians<sup>7</sup> and students<sup>8</sup> solving one single case out of 4 possible situations suggested that early hypothesis generation provided a structure to guide

physicians' acquisition of key clinical data. Further studies<sup>9,10</sup> also suggested that some behaviors in data collection, such as detailed inquiry about the chief complaint and frequent summarization of the collected information, were associated with better diagnostic outcomes. Despite the existing evidence, faulty data collection and interpretation are still important sources of errors<sup>11</sup> and many clinician educators still reward thoroughness of data collection rather than relevance dictated by initial diagnostic hypotheses. This study aims to confirm these principles with a larger set of cases from different organ systems and to determine the respective contribution of clinical experience and specific features of data collection and interpretation to explain diagnostic accuracy.

## METHODS

### Subjects and Research Design

We asked the 10 experienced General Internists heavily involved in teaching in our service to volunteer for our study. Six of them accepted, according to their time constraints. We then recruited second-year residents and senior medical students during successive residency and clerkship rotations in our service, until we obtained 6 participants in each group. All subjects worked up the same 7 chief complaints with a standardized patient, thus producing a total amount of 42 encounters for each group of clinical experience, a sample size estimated adequate in terms of power and feasibility. No specific review was required in our institution for this study.

We used charts of real patients to create 7 case scripts portrayed by a standardized patient (SP). Their chief complaints were: (1) heavy sensation in the abdomen, (2) cough, (3) weight loss, (4) headache, (5) diarrhea, (6) lower limb edema, and (7) arthritis. The diagnoses of these common cases relied mainly on history and physical examination.

All subjects encountered the 7 cases in the same order without time limitation. At the end of each encounter they provided their final working diagnosis. The encounters were audiotaped and immediately replayed for a thinking-aloud stimulated recall,<sup>1</sup> during which the subjects indicated the purposes underlying their data collection. These comments were audiotaped and retranscribed for analyses. Two previously trained investigators evaluated and tallied the characteristics of each encounter. Their interrater correlation ranged from 0.83 to 0.98.

---

This paper was presented in part at the annual conference of the Swiss Society of General Internal Medicine, Basel, 2005.

Address correspondence and requests for reprints to Dr. Nendaz: Department of Internal Medicine, Service of General Internal Medicine, Geneva University Hospitals, 1211 Geneva 14, Switzerland (e-mail: Mathieu.Nendaz@hcuge.ch).

---

Manuscript received December 3, 2005

Initial editorial decision January 18, 2006

Final acceptance June 19, 2006

Table 1. Characteristics of The Encounters, According to Clinical Experience\*

	Experts 41 encounters	Residents 42 encounters	Students 42 encounters	Experience effect (P*)	Case effect (P*)
Information collected					
Encounter duration, mean/case (minutes)	15.2 (13.8 to 16.7)	19.0 (18.0 to 19.9)	21.4 (19.6 to 23.3)	.03	.90
Unique findings collected, mean N/case	61 (56 to 67)	77 (72 to 83)	73 (67 to 79)	.19	.62
Relevance score <sup>†</sup> of unique findings, mean/case	0.60 (0.57 to 0.62)	0.41 (0.40 to 0.42)	0.43 (0.41 to 0.44)	<.0001	.68
Key questions <sup>‡</sup> , mean N/case	9 (8 to 10)	8 (7 to 8)	7 (6 to 8)	<.0001	<.0001
Summary occurrences, mean N/case	1.93 (1.63 to 2.22)	1.38 (1.07 to 1.69)	1.17 (.88 to 1.46)	.11	.59
Systems explored <sup>‡</sup>					
Body systems explored; mean N/case	7.4 (6.9 to 8.0)	7.4 (6.8 to 7.9)	6.8 (6.2 to 7.4)	.12	.21
Lines of inquiry, history, mean N/case	14 (12 to 16)	18 (16 to 20)	17 (15 to 20)	.41	.77
Diagnostic hypotheses					
Diagnostic hypotheses evaluated; mean N/case	14 (12 to 15)	16 (15 to 18)	16 (14 to 17)	.41	.04
Relevance of diagnostic hypotheses <sup>†</sup> , mean/case	0.69 (0.66 to 0.72)	0.49 (0.46 to 0.52)	0.49 (0.46 to 0.52)	<.001	.83
Findings collected until final diagnosis first generated, mean N/case	9.8 (7 to 12)	24 (16 to 32)	23 (15 to 32)	.008	.03
Final decisions					
Unique decisions made, mean N/case	7 (6 to 8)	8 (7 to 9)	8 (7 to 9)	.36	.005
Relevance of distinct decisions <sup>†</sup> , mean/case	0.69 (0.64 to 0.73)	0.42 (0.37 to 0.47)	0.52 (0.47 to 0.56)	<.001	.21

\*ANOVA with subjects nested within experience levels and repeated measures for cases. Numbers in brackets denote 95% confidence intervals.

<sup>†</sup>Relevance of information collected, diagnostic hypotheses generated, or decisions made, is their level of concordance (from 0, 0% concordance to 1, 100% concordance) among experts reaching the correct diagnoses. Key questions, decisions, or diagnostic hypotheses are those elicited by all members of this reference group.

<sup>‡</sup>Examples of body systems: respiratory, neurological. One line of inquiry is a sequence of consecutive questions evaluating the same diagnostic hypothesis.

## Outcome Variables and Data Analyses

We analyzed 125 encounters, 1 encounter being not recorded for technical problems. For each encounter, we determined the diagnostic accuracy (binary variable, based on the actual patient's diagnosis), the amount, relevance, and sequence of the information collected, the organ systems explored, the diagnostic hypotheses evaluated, and the management decisions made. Because there is no gold standard to work up specific cases, we used the level of concordance among experts with correct final diagnoses to determine the relevance of the information collected and the diagnostic hypotheses generated.<sup>12-15</sup> Each piece of information and diagnostic hypothesis received a relevance weight ranging from 0 (0% concordance) to 1 (100% concordance). Key information or hypotheses were those elicited by all experts (100% concordance).

We built an ANOVA model in which the unit of analysis was the encounter, i.e., the product of subjects (18) by cases (7), subjects being nested within 3 experience levels. We analyzed the effects of clinical experience on the variables listed in Table 1, with the 7 cases as repeated measures. We also tested interactions between cases and experience levels.

We determined the features of the data collection process predicting diagnostic accuracy by univariate, bivariate (correction for clinical experience), and multiple logistic regression models (corrected for all collected data). Standard errors and 95% confidence intervals (CI) were adjusted for intragroup correlation, thus taking into account the fact that the same subjects assessed many cases. All analyses were performed using the Stata<sup>®</sup> statistical software (release 9.1, Stata Corp., College Station, TX).

## RESULTS

The characteristics of the encounters differed according to the subjects' levels of clinical experience (Table 1). Overall, experts differed more from residents and students than did residents

from students. Compared with experienced physicians, younger doctors collected less relevant data; evaluated less relevant diagnostic hypotheses; evaluated the final correct diagnosis later during the encounter; and made decisions of lower relevance. No interaction between case and level of experience was significant. The proportion of cases diagnosed correctly was, respectively, 81% (95% CI 66 to 90), 45% (95% CI 31 to 60), and 36% (95% CI 23 to 51) for the experts, residents, and students ( $P < .001$ ).

The following variables significantly predicted diagnostic accuracy in the univariate logistic regression: higher level of clinical experience (odds ratio [OR] 7.43, 95% CI 2.17 to 25.41), collection of key information (OR 1.23, 1.09 to 1.39), summarization of available information (OR 1.50, 1.00 to 2.27), generation of the correct diagnosis at least once during the encounter (OR 15.45, 1.87 to 127.83), evaluation of the correct diagnosis within the first 10 questions asked (OR 28.29, 3.33 to 239.95), and evaluation of key diagnostic hypotheses during the encounter (OR 2.54, 1.54 to 4.18).

After correction for clinical experience (Table 2), frequent summarization of information was no longer significant and the total number of diagnostic hypotheses evaluated during the encounters became a significant predictor. The number of key diagnostic hypotheses remained the most significant variable, even with the conservative Bonferroni's correction for multiple comparisons.<sup>17</sup>

With multiple logistic regression analysis, clinical experience at the student level (OR 0.24, 0.07 to 0.83), evaluation of key diagnostic hypotheses during the encounters (OR 3.12, 1.55 to 6.25), and the late evaluation of the correct diagnosis (OR 0.97, 0.94 to 0.99) remained significant independent predictors of diagnostic accuracy (40% of the variance explained).

## DISCUSSION

In this study, several characteristics in data collection and interpretation predicted diagnostic accuracy beyond the accu-

**Table 2. Bivariate Logistic Regression Analysis Corrected for Experience for the Prediction of Complete and Correct Diagnosis, with 95% Confidence Intervals Adjusted for Intrasubject Correlation**

	Odds ratio	95% CI	P*
Mean number of key questions asked by case <sup>†</sup>	1.19	1.04 to 1.36	.01
Mean number of lines of inquiry by case <sup>‡</sup>	1.05	1.01 to 1.11	.03
Mean number of diagnostic hypotheses evaluated by case	1.08	1.01 to 1.16	.02
Mean number of key diagnostic hypotheses evaluated by case	2.22	1.34 to 3.67	.002
Correct diagnostic hypothesis evaluated at least once during the encounter	15.17	1.05 to 219.6	.04
Correct diagnostic hypothesis generated within the first 10 questions asked	24.35	2.66 to 222.50	.005

\*If Bonferroni's correction for multiple comparisons is applied, the significance threshold becomes 0.005.

<sup>†</sup>Key questions or diagnostic hypotheses are those elicited by all members of the reference group of experts reaching the correct diagnoses.

<sup>‡</sup>One line of inquiry is a sequence of consecutive questions evaluating the same diagnostic hypothesis.

CI, confidence interval.

mulated years of practice, among which the most important were the collection of key information, the evaluation of relevant diagnostic hypotheses and the generation of the correct diagnosis within the first 10 questions asked during the encounter. This highlights the crucial importance of an early evaluation of relevant diagnostic hypotheses during the work-up to diagnose successfully a case, as it drives the subsequent collection of relevant information. Our results on several cases in various domains of internal medicine expand previous research that already showed these relationships with few cases<sup>1,7,9</sup> from specific specialties (e.g., neurology) or cases relying on visual cues.<sup>3-6</sup> In addition, some previous works relied on written clinical vignettes rather than higher-fidelity simulation allowing for an open-ended inquiry (e.g., standardized patients), a condition known to alter clinical reasoning because the information is immediately provided rather than progressively collected by the subject.<sup>16,18</sup> Our data also give an additional insight into the role of clinical experience. While a focused data collection and frequent summarizations of the collected clinical data are more a trait of a higher level of training than a necessary condition of diagnostic success, the exploration of a larger number of diagnostic hypotheses becomes an important clue for successful younger subjects. More than accumulated years of practice, previous exposure to similar cases may thus represent an important determinant of diagnostic success, as also suggested by the tiny differences observed between the characteristics of residents and students.

Many of these principles have already been suggested by medical educators but their internalization by clinician-educators remains difficult in practice. By actualizing them, our data reinforce the goals medical trainers should strive to attain with their trainees and give credence to teaching activities fostering the exploration of diagnostic hypotheses related to the patient's complaint and their use to frame further data collection.<sup>19</sup> Whatever the teaching strategy, it should favor the simultaneous acquisition of knowledge and process to remain optimal.<sup>20</sup> Our results also support teaching programs that offer early and systematic approach to a variety of practical cases and do not merely rely on a random and uneven exposure.

This study has some limitations restricting the generalization of the results. First, it was conducted in a single institution with volunteers. The subjects were, therefore, possibly more motivated than those who declined participation, although this selection bias would have rather reduced the differences we observed among groups of different levels of clinical experience. Second, although the standardization of

the setting increases reliability, it may hinder the natural reasoning the same physicians would have when facing a real patient in a natural setting.

In conclusion, some characteristics of clinical data collection are related to diagnostic accuracy beyond traits more directly related to clinical experience. Medical educators should consider them as training goals for learners in clinical environments and reinforce the importance of using an early and wide exploration of diagnostic hypotheses to frame clinical data collection. This implies a more explicit role modeling of clinical reasoning and the abandonment of the still prevailing sense that exhaustive data collection is the privileged way to diagnostic success.

*We thank the faculty members, residents, and students who so willingly participated in this study.*

*Funding sources: Swiss National Science Foundation, Grant no. 3200B0-102265/1 and Elie Safra Foundation, Geneva, Switzerland.*

## REFERENCES

1. **Elstein AS, Shulman LS, Sprafka SA.** Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, MA: Harvard University Press; 1978.
2. **Norman GR.** Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39:418-27.
3. **Norman GR, Brooks LR, Cunnington JPW, Shali V, Marriott M, Regehr G.** Expert-novice differences in the use of history and visual information from patients. *Acad Med.* 1996;71:62-4.
4. **Brooks LR, LeBlanc VR, Norman GR.** On the difficulty of noticing obvious features in patient appearance. *Psychol Sci.* 2000;11:112-7.
5. **Leblanc VR, Brooks LR, Norman GR.** Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med.* 2002;77:S67-S69.
6. **Leblanc VR, Norman GR, Brooks LR.** Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med.* 2001;76:S18-S20.
7. **Barrows HS, Norman GR, Neufeld VR, Feightner JW.** The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med.* 1982;5:49-55.
8. **Neufeld VR, Norman GR, Barrows HS, Feightner JW.** Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *Med Educ.* 1981;15:315-22.
9. **Hasnain M, Bordage G, Connell KJ, Sinacore JM.** History-taking behaviors associated with diagnostic competence of clerks: an exploratory study. *Acad Med.* 2001;76:S14-S17.
10. **Nendaz MR, Gut AM, Perrier A, et al.** Common strategies in clinical data collection displayed by experienced clinician-teachers in internal medicine. *Med Teach.* 2005;27:415-21.
11. **Bordage G.** Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med.* 1999;74:S138-S143.

12. **Norman GR, Barrows HS, Feightner JW, Neufeld VR.** Measuring the outcome of clinical problem-solving. *Ann Conf Res Med Educ.* 1977;16:311-6.
13. **Charlin B, Desaulniers M, Gagnon R, Blouin D, van der Vleuten C.** Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. *Teach Learn Med.* 2002;14:150-6.
14. **Nendaz MR, Gut AM, Perrier A, et al.** Degree of concurrency among experts in data collection and diagnostic hypothesis generation during clinical encounters. *Med Educ.* 2004;38:25-31.
15. **Norcini J, Shea J, Day S.** The use of the aggregate scoring for a recertification examination. *Eval Health Prof.* 1990;13:241-51.
16. **Nendaz M, Raetzo M, Junod A, Vu.** Teaching diagnostic skills: clinical vignettes or chief complaints? *Adv Health Sci Educ Theory Pract.* 2000;5:3-10.
17. **Perneger TV.** What's wrong with Bonferroni adjustments. *BMJ.* 1998;316:1236-8.
18. **Gruppen LD, Wolf FM, Billi JE.** Information gathering and integration as sources of error in diagnostic decision making. *Med Decis Making.* 1991;11:233-9.
19. **Kassirer JP.** Teaching clinical medicine by iterative hypothesis testing. Let's preach what we practice. *N Engl J Med.* 1983;309:921-3.
20. **Eva KW.** What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39:98-106.