ORIGINAL RESEARCH

Accuracy of diagnoses predicted from a simple patient questionnaire stratified by the duration of general ambulatory training: an observational study

Takanori Uehara[1,2]
Masatomi Ikusaka[1]
Yoshiyuki Ohira[1]
Mitsuyasu Ohta[1,2]
Kazutaka Noda[1]
Tomoko Tsukamoto[1]
Toshihiko Takada[1]
Masahito Miyahara[1]

[1]Department of General Medicine, Chiba University Hospital, [2]Division of Rotated Collaboration Systems for Local Healthcare, Graduate School of Medicine, Chiba University, Chiba, Japan

Purpose: To compare the diagnostic accuracy of diseases predicted from patient responses to a simple questionnaire completed prior to examination by doctors with different levels of ambulatory training in general medicine.

Participants and methods: Before patient examination, five trained physicians, four short-term-trained residents, and four untrained residents examined patient responses to a simple questionnaire and then indicated, in rank order according to their subjective confidence level, the diseases they predicted. Final diagnosis was subsequently determined from hospital records by mentor physicians 3 months after the first patient visit. Predicted diseases and final diagnoses were codified using the International Classification of Diseases version 10. A "correct" diagnosis was one where the predicted disease matched the final diagnosis code.

Results: A total of 148 patient questionnaires were evaluated. The Herfindahl index was 0.024, indicating a high degree of diversity in final diagnoses. The proportion of correct diagnoses was high in the trained group (96 of 148, 65%; residual analysis, 4.4) and low in the untrained group (56 of 148, 38%; residual analysis, −3.6) (χ^2=22.27, $P<0.001$). In cases of correct diagnosis, the cumulative number of correct diagnoses showed almost no improvement, even when doctors in the three groups predicted ≥4 diseases.

Conclusion: Doctors who completed ambulatory training in general medicine while treating a diverse range of diseases accurately predicted diagnosis in 65% of cases from limited written information provided by a simple patient questionnaire, which proved useful for diagnosis. The study also suggests that up to three differential diagnoses are appropriate for diagnostic prediction, while ≥4 differential diagnoses barely improved the diagnostic accuracy, regardless of doctors' competence in general medicine. If doctors can become able to predict the final diagnosis from limited information, the correct diagnostic outcome may improve and save further consultation hours.

Keywords: clinical reasoning, diagnostic accuracy, diagnostic reasoning, general medicine, Herfindahl index, predict disease

Introduction

Research on clinical reasoning started in the 1960s. Doctors have been shown to engage in clinical reasoning through backward analytical reasoning, represented by the hypothetico-deductive method,[1] as well as forward nonanalytical reasoning, represented by pattern recognition.[2] In recent years, an approach combining both forms of reasoning has been suggested,[3,4] with Eva et al[5] reporting that utilizing a combined reasoning approach improved the accuracy of diagnostic reasoning. Expert physicians, when faced

Correspondence: Takanori Uehara
Department of General Medicine,
Chiba University Hospital, 1-8-1 Inohana,
Chuo-ku, Chiba 260-8677, Japan
Tel +81 43 222 7171
Fax +81 43 224 4758
Email takanori.ue@nifty.com

with an undifferentiated diagnostic problem, are reported to reduce uncertainty by generating one or more diagnostic hypotheses and then searching for additional information to confirm or refute one or more of the hypotheses in order to reach a final diagnosis.[6] Thus, the generation of diagnostic hypotheses plays an important role in problem-solving.[7]

Elstein et al[1] reported that regardless of competency levels, medical students and physicians generate 4±1 hypotheses when engaged in diagnostic reasoning at any one time, and Barrows et al[8] found that in a study using a standardized simulated patient that early generation of the correct diagnosis correlates significantly with the correct diagnostic outcome. Within a few minutes of an encounter with a patient, experts are likely to generate several hypotheses based on limited medical history information and engage in diagnostic reasoning.[9] On the other hand, the number of hypotheses generated showed no correlation with the proportion of correct diagnostic outcome.[8] Because short-term memory plays an important role in generating hypotheses[10] and only a limited number of diseases can be predicted in diagnostic reasoning, the quality of the differential diagnoses made is important.[8]

To the best of our knowledge, no studies have compared the diagnostic accuracy of predicting diseases made on the basis of patient responses to a simple patient questionnaire or have investigated the appropriate number of differential diagnoses in actual clinical encounters according to the duration of ambulatory training at a department of general medicine. Therefore, in this study we examined the diagnostic accuracy of diseases predicted on the basis of patients' written responses to a simple questionnaire at an early stage in the patient-examination process. We then compared the diagnostic accuracy of doctors and the number of differential diagnoses with different levels of ambulatory training in general medicine. We also prepared a rank-order list of differential diagnoses made by doctors according to their own subjective levels of confidence in order to investigate the appropriate number of predicted diseases in diagnostic reasoning. Finally, we examined the usefulness of a simple questionnaire in an actual clinical setting.

Materials and methods
Setting
This study was conducted in a hospital affiliated with Chiba University School of Medicine, located in the center of Chiba City, which is home to a population of 950,000 people and located 40 kilometers from the capital, Tokyo, in Japan. The hospital is a tertiary medical facility that treats approximately 2,000 patients daily. The Department of General Medicine affiliated with Chiba University School of Medicine examines adult patients without a referral or those with an unknown diagnosis with a referral from a department within the hospital or from other hospitals and clinics. Examinations are conducted from 8.30 am to 5 pm Monday through Friday, excluding national holidays. The department's outpatient program for doctors operates similarly to the "resident-as-teacher program,"[11] with residents who have received short-term training (short-term-trained group) supervising a group of residents who have not yet received any training (untrained group), and physicians who have received training (trained group) supervising the short-term and untrained groups. On average, doctors conduct examinations for two new patients and three returning patients each day, 4 days a week. The academic year in Japan begins in April and ends the following March.

Participants and study design
The study was conducted with all physicians involved in providing treatment at the time of the study (from April to May 2010), at a general medicine outpatient department that uses a questionnaire, completed in writing by patients at their first visit. We assigned participants to one of three groups: physicians who had completed general ambulatory training (3 years' duration) in the department (trained group); residents who had undergone a short period (1 year's duration) of general ambulatory training in the department (short-term-trained group); and residents who had not started the general ambulatory training in the department (untrained group). The patient questionnaire asked questions about age, sex, chief complaint, duration of symptoms, history of hospital treatment, past medical history, allergies to medication, preferences for smoking and drinking, pregnancy, and breastfeeding. The space for patients to provide each response was limited to within one line on B5 paper. We routinely use this simple questionnaire at patients' first visit to the department (see Supplementary material).

In this study, teams of three physicians – one physician each from the trained, short-term-trained, and untrained groups – were formed randomly to participate in the routine examination and training carried out in the department, and patients were randomly assigned. If the untrained group was already in the process of examining a patient and was unable to take on a new patient, then the short-term-trained group or the trained group examined the patient without the patient first being subjected to examination by the untrained group. We analyzed only those cases where all three doctors in a group examined the patients.

We conducted the study during daily clinical activities. We gave the simple questionnaire completed by each patient

to all three doctors stratified by ambulatory training in a team, and asked them to indicate on the survey sheet, in rank order according to their own subjective level of confidence, the differential diagnoses (hereafter "predicted diseases") they generated. We codified the predicted diseases and final diagnoses using the International Classification of Diseases version 10.

At 3 months after the first visit, mentor physicians who were blinded to the responses on the questionnaire sheets and were not involved in the patient examinations made the final diagnoses based on the patients' medical records. The Research Ethics Committee of Chiba University School of Medicine approved the study protocol. Participants gave informed consent prior to their participation.

Statistical analysis

First, we examined the degree of diversity of final diagnoses using the Herfindahl index (HI),[12] because a skewed distribution of diseases would affect the study outcomes. We obtained the HI by summing the squares of the share of each diagnostic category used: a score of 1 means only one diagnostic category is used, whereas if all categories are used equally, the score approaches 0.[12] Second, we considered a "correct" diagnosis to be a match between a predicted disease and the final diagnosis code. A table of correct diagnoses among the three groups was created and examined using a χ^2 test. Factors in sections of the cross table that showed differences were examined by residual analysis. Third, we compared the number of predicted diseases between the groups using the Mann–Whitney U test and multiple comparisons with Bonferroni correction. Fourth, for cases of correct diagnosis, we analyzed the cumulative proportions of correct diagnoses by confidence level for each group.

All statistical analysis was performed using SPSS for Windows version 20.0 (IBM Corporation, Armonk, NY, USA), with significance set at $P<0.05$. In residual analysis, an absolute value of 1.96 for adjusted residual errors was considered to be a cell that would disturb comparability between the groups. Significance on the Mann–Whitney U test and multiple comparisons was set at $P<0.05/3=0.017$.

Results

Five physicians participated in the trained group, four in the short-term-trained group, and four in the untrained group (Table 1). During the study period, we included 156 cases, but after excluding eight (5.1%) due to uncompleted questionnaires or because the patients were shown to be asymptomatic by further examination of irregular findings, this left 148 fully completed questionnaires (response rate 94.9%) for analysis.

Patient characteristics were as follows: 63 men (43%), 85 women (57%), mean age 50 years, and 60 patients (41%) referred to us. Final diagnoses involved 17 areas of the ICD-10 and 80 codes, and the HI was 0.024.

The proportion of correct diagnoses was 65% (96 of 148) for the trained group, 47% (70 of 148) for the short-term-trained group, and 38% (56 of 148) for the untrained group, yielding significant differences between the three groups (χ^2=22.27, $P<0.001$). Residual analysis revealed the proportion of correct diagnoses was high in the trained group (residual analysis, 4.4) and low in the untrained group (residual analysis, −3.6).

The median number (25th and 75th percentiles: Q1 and Q3, respectively) of predicted diseases was three (two and three) for the trained group, four (three and five) for the short-term-trained group, and three (two and four) for the untrained group. Multiple comparisons revealed that compared to the short-term-trained group, the untrained group generated significantly fewer predicted diseases, as did the trained group (P=0.001 and $P<0.001$, respectively). In contrast, there was no significant difference between the untrained and trained groups (P=0.037).

The cumulative number of correct diagnoses by confidence level in cases of correct diagnoses for each group barely improved, even when the physicians in the three groups predicted ≥4 diseases (Figure 1).

Discussion

The final diagnoses of the 148 cases analyzed in this study were quite diverse, as shown by the HI of 0.024, given that the HI for the US National Ambulatory Medical Care

Table 1 Demographics of the three groups of participating doctors (n=13) at a general medicine outpatient department

Group	Participant doctors, n	Male/female	Mean age (years)	Duration of ambulatory training at a general medicine department (years)
Trained group	5	5/0	32	3–5
Short-term-trained group	4	4/0	28	1
Untrained group	4	2/2	26	0

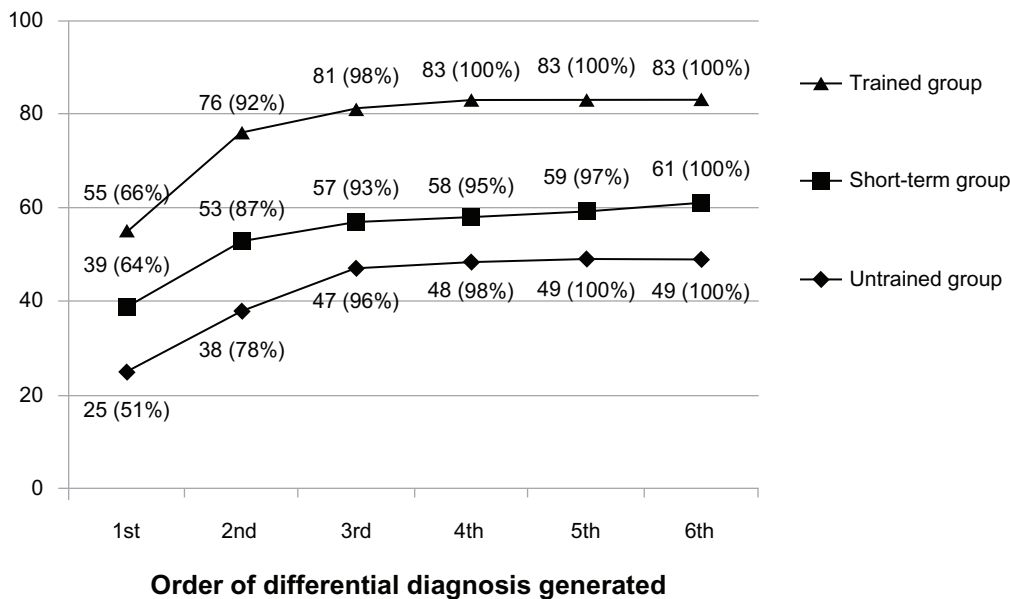**Cumulative no of accurate diagnoses**

Figure 1 Cumulative number of accurate diagnoses by rank order of certainty in cases of accurate diagnoses in each group.

Surveys was 0.19 for general practice, 0.15 for family practice, 0.53 for cardiology, and 0.34 for gastroenterology.[13] In addition to specialized areas, such as gastroenterology and cardiology, scores in this study were even lower than in general practice and family practice. In the Japanese medical system, visiting a university hospital based on an individual's free will is guaranteed. Thus, not only referred patients having diseases with a low base rate but also patients with common diseases visit the Department of General Medicine, making for a high degree of diversity in the diseases seen at the department.

The trained group accurately predicted diagnosis in approximately 60% of all cases from the limited written information provided by the simple patient questionnaire that was conducted in a general outpatient facility treating a wide variety of diseases. The diagnostic accuracy was high in the trained group and low in the untrained group. Gruppen et al's[14] study of medical students' problems with patient management found that compared with students who did not include the correct diagnosis among the differential diagnoses, students who did include it (either as a primary or a secondary candidate) after the chief complaint were 3.5 times more likely to reach the correct diagnosis in cases of rheumatoid arthritis and 8.7 times more likely in cases of systemic lupus erythematosus. These results suggest that the trained group will form a more accurate final diagnosis after taking the history and completing physical and laboratory

investigations, because they can generate more accurate diagnostic hypotheses based on only the limited written information provided by a simple patient questionnaire.

In the present study, all three groups predicted around three possible hypotheses on average. This is largely consistent with Elstein et al's[1] findings that medical students and physicians generate 4±1 diagnostic hypotheses as they reason. Here, we found that the number of suspected diseases increased from the untrained group to the short-term-trained group, and then decreased in the trained group. In the untrained group, even though the residents possessed medical knowledge, their reasoning method was immature, they could not link this to the patient's chief complaint, and therefore they predicted fewer diseases with low diagnostic accuracy. In the short-term-trained group, the number of diseases that they could link to the chief complaint increased, making it possible to produce a higher number of possible diseases. The results of the trained group suggest that their reasoning method had matured and that a competence level existed in which diagnostic links were sufficiently refined not to predict a large number of diseases.

The cumulative proportions of correct diagnoses ranked by confidence level in cases of correct diagnosis for each group (Figure 1) revealed that even if ≥4 diseases were predicted, the proportion of accurate diagnoses barely improved. This suggests that up to three differential diagnoses might be suitable for predicting diagnosis. Thus, in cases where

doctors think that the final diagnosis might not match one of their top three predicted diseases at the time of diagnostic reasoning, their diagnostic accuracy will likely not improve by generating a fourth or more predictions. Rather, additional information or a different approach to diagnostic reasoning may be required.

In this study, doctors generated diagnostic hypotheses based solely on written information provided by patients prior to examination. In the same way as in other medical disciplines, we expected that diagnostic reasoning in general outpatient services would generate predictions from a small number of keywords searched in long-term memory. Mental representation (eg, semantic qualifiers, scripts, schema, and exemplars) is formed from deliberate practice with multiple examples, with feedback facilitating the acquisition of expertise in predicting a diagnosis and with experience gained by forming final diagnoses in an increasing number of cases also being critical to developing competence.[9] Although our study did not make doctors' decision-making process clear, we assume that physicians in the trained group broke down bits of information and used mental representations to accurately predict final diagnoses. Repetition of this kind of analytical thinking may result in nonanalytical thinking, such as pattern recognition, and enhance diagnostic reasoning competence.

The present findings revealed that final diagnoses could be predicted from limited information derived from a simple patient questionnaire in approximately 60% of cases when the doctors had expertise in diagnostic reasoning. Accurate prediction of final diagnosis during the early phase of an examination will shorten the time needed for examination while improving efficiency. In contrast to the established examinations for hospitalized patients, further education is needed for diagnosing outpatients on the basis of limited information, due to limited time available for diagnosis.

Research is now needed on actual methods to enhance such expertise. It appears that comprehensive clinical information is not needed for diagnostic reasoning. Rather, by using information obtained by questionnaire, it seems possible that a method of rank-ordering possible diagnoses from limited patient information according to the doctor's level of confidence could be applied to medical education and professional physician development.

Limitations

As the present research was conducted at only one facility and the number of doctors who participated in the study was small, further study should be conducted at multiple facilities to produce generalizable results. We could not evaluate the proportion of accurate diagnoses for each item on the questionnaire. Although the simple survey suggested that patient age, sex, and chief complaints are useful, further examination in this area is required in future studies.

Conclusion

Doctors who had completed an ambulatory training program at a general medicine outpatient facility treating diverse diseases accurately predicted diagnoses in 65% of cases on the basis of limited written information provided by a simple patient questionnaire. Thus, a clinical questionnaire is useful to doctors when making a definitive diagnosis, and accurate prediction of final diagnoses based on limited sources of information can shorten the time needed for examination and improve diagnostic accuracy. Increased experience resulted in more mature inference methods and a lower number of predicted diseases. Furthermore, the results suggest that up to three differential diagnoses are appropriate in predicting diseases, while ≥ 4 differential diseases barely improved diagnostic accuracy, regardless of doctors' competency levels.

Disclosure

References

1. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge (MA): Harvard University Press; 1978.
2. Groen GJ, Patel VL. Medical problem-solving: some questionable assumptions. *Med Educ*. 1985;19:95–100.
3. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84:1022–1028.
4. Boreham NC. The dangerous practice of thinking. *Med Educ*. 1994;28:172–179.
5. Eva KW, Hatala RM, Leblanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ*. 2007;41:1152–1158.
6. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ*. 2007;41:1140–1145.
7. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002;324:729–732.
8. 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.

9. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418–427.

10. Elstein AS. Thinking about diagnostic thinking: a 30-year perspective. *Adv Health Sci Educ Theory Pract*. 2009;14 Suppl 7:7–18.

11. Hill AG, Yu TC, Barrow M, Hattie J. A systematic review of resident-as-teacher programmes. *Med Educ*. 2009;43:1129–1140.

12. Franks P. Why do physicians vary so widely in their referral rates? *J Gen Intern Med*. 2000;15:163–168.

13. Franks P, Clancy CM, Nutting PA. Defining primary care. Empirical analysis of the National Ambulatory Medical Care Survey. *Med Care*. 1997;35:655–668.

14. Gruppen LD, Palchik NS, Wolf FM, Laing TJ, Oh MS, Davis WK. Medical student use of history and physical information in diagnostic reasoning. *Arthritis Care Res*. 1993;6:64–70.

Supplementary material
Patient questionnaire used for predictive diagnosis

Please answer the following questions to the best of your knowledge, as this information will be used in diagnosis and treatment.

Please answer in regard to the symptoms that brought you here today.
1. Please describe your symptoms (or illness).
2. When did you start to experience these symptoms?
3. Have you received treatment for these symptoms? (Please write down any over-the-counter medications you take for these.)

Hospital/clinic name

Please answer in regard to any previously experienced diseases.

Are you undergoing treatment at present?   Yes/No

If yes, please write the name of the disease.

1._____ 2._____ 3._____

Have you ever been seriously ill?   Yes/No

If yes, please write the name of the disease and the time period.

1._____ 2._____ 3._____

Please answer in regard to allergies to medicine.

Have you ever had an allergic reaction to medicine?   Yes/No

If you answered yes, what was the name of the medication and what were your symptoms?

Please answer in regard to your lifestyle habits.

Smoking   (not at all/smoke _____ cigarettes a day)

Drinking alcohol   (not at all/drink alcohol)

Please answer if you are female.

Is there a chance that you are pregnant, or are you currently breastfeeding?   Yes/No