# Effects of Differential Diagnosis Checklist and General De-Biasing Checklist on Diagnostic Performance in Comparison to Intuitive Diagnosis

**Authors:**

Taro Shimizu, MD, MPH; Kentaro Matsumoto, MD; Yasuharu Tokuda, MD, MPH

**Affiliations:**

(TS, YT) Mito Medical Center, University of Tsukuba, Mito City, Ibaraki, Japan; (KM) National Medical Clinic, Tokyo, Japan

**Corresponding author:**

Taro Shimizu, MD, MPH
Mito Medical Center, University of Tsukuba,
3-2-7 Miya-machi, Mito City, Ibaraki, 310-0015 Japan
Phone: 81-29-231-2371 FAX 81-29-221-5137
Email: shimizutaro7@gmail.com

**Abstract**

**OBJECTIVE:**

We aimed to evaluate diagnostic performance by intuitive process (system 1) and analytic process (system 2) among medical students. In this study, as aiding tools of analytic process, we used a differential diagnosis checklist (DDXC) and a general de-biasing checklist (GDBC).

**METHODS:**

We enrolled 188 Japanese medical students (4th to 6th grades) who were divided into 2 groups and assigned the same five cases scenarios. Group 1 students (n=91) were instructed to provide the 3 most likely diagnoses immediately after reading the case scenarios (intuitive diagnosis), then after reading GDBC (diagnosis by GDBC), and finally, after reading DDXC (diagnosis by DDXC). Conversely in order, group 2 students (n=97) were instructed to provide the 3 most likely intuitive diagnoses, then diagnosis by DDXC, and finally, diagnosis by GDBC.

**RESULTS:**

Among the group 1, while there was no significant difference of total scores between intuitive diagnosis (8.25) and diagnosis by GDBC (8.42), there was significant difference of total scores (P = 0.01 by ANOVA) between intuitive diagnosis and diagnosis by DDXC (8.77). Among the group 2, there was significant difference of total scores (P = 0.001 by ANOVA) between intuitive diagnosis (7.21) and diagnosis by DDXC (7.96), although there was no significant difference of total scores between diagnosis by DDXC and GDBC (8.12). Among the difficult case scenarios (cases 3, 4, 5), the proportions of correct diagnosis increased after reading DDXC in both groups. Among simple case scenarios (cases 1, 2), the proportions of correct diagnosis decreased after reading DDXC among the group 2.

**CONCLUSIONS:**

The use of DDXC may improve the diagnostic performance in difficult cases, while intuitive process may still be better for diagnosing simple cases. Further studies are needed for examining actual clinical usefulness of intuitive diagnosis as well as DDXC.

## Introduction

One of crucial factors in physicians' clinical performances is their diagnostic reasoning skill[1]. Recent studies suggest that actual processes of diagnostic reasoning can be classified into two mental processes referred to as the intuitive process (system 1) and analytical process (system 2)[14,26,27]. This "dual-processing" model implies that physicians are considered to subconsciously use both systems as the situation demands during their diagnostic reasoning. Intuitive process rapidly occurs by unconscious mental matching of a given clinical case to a prior experience based on heuristics[28]. Analytical process is slow but more comprehensive by consciously analyzing a clinical case based on probability theory, textbooks, and checklists[29].

A flaw of diagnostic reasoning can cause diagnostic errors that could be linked to higher morbidity and mortality in affected patients[2-7]. Multiple studies suggest that most flaws of diagnostic reasoning might derive from error of intuitive process (the theory of intuitive error), and that a variety of measures has been proposed to reduce the failure of intuitive process[1, 8, 9]. The candidate measures included reflective feedback, cognitive forcing training, electronic differential diagnosis generators, or checklists[10-25].

Recently, Ely, Graber, and Croskerry suggested that, based on the theory of intuitive error, using a general de-biasing checklist (GDBC) and a differential diagnosis checklist (DDXC) may be effective for reducing diagnostic errors from intuitive process, since these may aid physicians as the measures enhancing analytical process[30]. The GDBC were developed to prompts physicians to optimize their cognitive approach by providing a reproducible approach to diagnosis (Appendix 2)[30]. The DDXC is developed to help physicians avoid the failure to consider the correct diagnosis as a possibility, which is the most common cause of diagnostic error, and it is used to prompt physicians to consider a comprehensive list of causes for the complaints that commonly present diagnostic challenges[30].The checklists highlight diagnoses that should not be missed and those that are, in fact, commonly missed[30].

However, to our knowledge, no studies have been conducted for examining the effects of checklists, such as GDBC and DDXC, for enhancing analytic process on diagnostic performance. Thus, in this study, we evaluated the influence of these checklists on diagnostic performance by comparing diagnostic accuracy with intuitive diagnosis among medical students. We also examined the differential effectiveness of the checklists by the relative difficulty of case scenarios.

## Methods

We conducted a study of diagnostic quiz cases at several universities in Okinawa, Osaka, Nara, Tokyo, Toyama, and Ibaraki, Japan in August 2011 to January 2012. Our diagnostic quiz cases contained 5 challenging cases to medical diagnosis; case 1, acute coronary syndrome; case 2, subarachnoid hemorrhage; case 3, Fitz-Hugh-Curtis syndrome; case 4, aortic dissection; and, case 5, obutrator hernia (Appendix 1). These cases scenarios were developed from actual patients by a group of experienced teaching physicians (three associate professors of medicine) in Department of Medicine, Toho University School of Medicine, Tokyo, Japan.

The cases were arranged based on diagnostic difficulty estimated by the group, being the more difficult cases by the increasing case number (case 1, the least difficult; case 5, the most difficult) by the consensus of the group of experienced teaching physicians on the contents of scenarios. The case scenario review was conducted and the content validity was confirmed by an independent panel of the university (three professors of medicine).

Each case scenario was followed by 3 parentheses, in which participants were required to write down the first most, the second most and the third most likely diagnoses in order. First, as intuitive diagnosis, participants were asked to write three likely diagnoses by quickly reading the case scenarios (5 minutes after reading each case). Next, for group 1 students (n=91), as diagnosis by GDBC, after receiving and reading GDBC, the general de-biasing checklist developed by Ely, Graber, and Croskerry[30] (Appendix 2), they were also asked to write three likely diagnoses (5 minutes for each case). Third, as diagnosis by DDXC, after receiving and reading DDXC, the differential diagnosis checklist suggested by Ely et al[30] and developed by our investigators based on multiple textbooks of differential diagnosis (Appendix 3), they were again asked to write three likely diagnoses (5 minutes for each case). DDXC included highlights on "Do not miss diagnosis" and "Frequently missed diagnosis". Conversely, for group 2 students (n=97), they were asked to write three likely diagnoses (diagnosis by DDXC) after receiving and reading DDXC and after receiving and reading GDBC (diagnosis by GDBC).

We used existing medical school conferences for implementing the study. These conferences were the regular teaching conferences. All participants were Japanese and they read and signed informed consent before the study. Participants were assured of confidentiality and anonymity. We allocated medical students of Okinawa, Osaka, and Ibaraki into the group 1 and those of Nara, Tokyo, and Toyama into the group 2. The study was approved by the Institutional Review Board of Mito Medical Center, University of Tsukuba, Mito City, Ibaraki, Japan.

We analyzed the mean scores by allocating scores of score 3 as a correct diagnosis for the first most, score 2 for the second most, and score 1 for the third most likely diagnoses. We compared mean total scores between intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC, using repeated measures of ANOVA adjusted for gender and grade, and the comparisons were performed using Sidak's analysis. In addition, when participants indicated a correct diagnosis in any of the first most, the second most and the third most likely diagnoses, they were considered as correct diagnosis. Proportion (%) of correctness of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC were calculated. All statistical analysis was performed using SPSS-J 17.0 (Tokyo, Japan) and two-tailed p-value of less than 0.05 was considered as statistically significant.

**Results**

A total of 188 medical students (128 men and 60 women) agreed to participate in our study. They included 35 4th-grade, 77 5th-grade and 76 6th-grade students. Table 1 shows score distribution of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC by groups 1 and 2.

Table 2 reveals mean score of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC by the group 1. Among cases 3, 4 and 5, in which mean baseline scores were relatively lower than those of cases 1 and 2, the mean scores of diagnosis by DDXC were higher than those of intuitive diagnosis and diagnosis by GDBC. However, among cases 1 and 2 with relatively higher mean scores, the mean scores of diagnosis by DDXC were lower than those of intuitive diagnosis and diagnosis by GDBC. Repeated measures analysis of ANOVA adjusted for gender and grade showed there was significant difference of total score among intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC (P=0.003). By Sidak's analysis, total scores between intuitive diagnosis and diagnosis by GDBC was not significantly different (P=0.218), while total scores between intuitive diagnosis and diagnosis by DDXC was significantly different (P=0.01).

Table 2 also reveals mean score of intuitive diagnosis, diagnosis by DDXC, and diagnosis by GDBC by the group 2. Among cases 3, 4 and 5, in which mean baseline scores were relatively lower than those of cases 1 and 2, the mean scores of diagnosis by DDXC were higher than those of intuitive diagnosis. However, among cases 1 and 2 with relatively higher mean scores, the mean scores of diagnosis by DDXC were lower than those of intuitive diagnosis. Repeated measures analysis of ANOVA adjusted for gender and grade showed there was significant difference of total score among intuitive diagnosis, diagnosis by DDXC, and diagnosis by GDBC (P=0.001). By Sidak's analysis, total scores between intuitive diagnosis and diagnosis by DDXC was significantly different (P=0.001), while total scores between diagnosis by DDXC and diagnosis by GDBC was not significantly different (P=0.25).

Figure 1 shows proportions of correct diagnosis of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC by the group 1. Among cases 3, 4 and 5, in which the proportions were relatively lower than those of cases 1 and 2, the proportions of correct diagnosis by DDXC were greater than those of intuitive diagnosis and diagnosis by GDBC. However, among cases 1 and 2 with relatively greater proportions of correct diagnosis, the proportions of correct diagnosis by DDXC were lower than those of intuitive diagnosis and diagnosis by GDBC. For all cases combined, 60% of correct diagnosis was observed for intuitive diagnosis, 62% for diagnosis by GDBC, and 67% for diagnosis by DDXC.

Figure 1 also shows proportions of correct diagnosis of intuitive diagnosis, diagnosis by DDXC, and diagnosis by GDBC by the group 2. Among cases 3, 4 and 5, in which the proportions of correct diagnosis were relatively lower than those of cases 1 and 2, the proportions of diagnosis by DDXC were greater than those of intuitive diagnosis, and only minimum difference of the proportions were noted between diagnosis by DDXC and GDBC. However, among cases 1 and 2 with relatively greater proportions of correct diagnosis, the proportions of diagnosis by DDXC were actually lower than those of intuitive diagnosis, and only minimum difference of the proportions were noted between diagnosis by DDXC and GDBC. For all cases combined, 53% of correct diagnosis was observed for intuitive diagnosis, 59% for diagnosis by DDXC, and 60% for diagnosis by GDBC.

**Discussion**

Our results suggest that the use of differential diagnosis checklists improved diagnostic performance, but that the use of a general de-biasing checklist did not. The effectiveness of the differential diagnosis checklist was not affected by the prior use of the proposed general checklist. Although total scores in all cases combined were significantly greater in reading a differential diagnosis checklist than without checklist, there was difference of the effects of a differential diagnosis checklist between difficult cases and less difficult cases. The use of a differential diagnosis checklist led to better diagnostic performance among difficult cases. However, it might lead to reduced performance among simple cases. To our knowledge, this is the first study to show that differential diagnosis checklists worked effectively for improving diagnostic performance.

Our results suggest that general de-biasing checklist did not work as a tool for better diagnostic performance. There are several reasons that could explain these results. First, each item of the checklist was not designed to derive any specific differential diagnosis, while it instructs overall approach to make differential diagnoses. Hence this checklist may work little in each specific clinical scenario. Second, because the participants of this study were all medical students, who were novice and had little clinical experience, the use of general de-biasing checklist might not have contributed to the baseline level with lack of heuristic thinking among students.

Based on the different levels of difficulty among case scenarios, there was a trend suggesting that the proportions of correct diagnoses after reading differential diagnosis checklists were reduced in cases with easier levels. According to the dual processing theory, analytical process including a checklist, works better for difficult/challenging cases and intuitive process may be more effective for solving easy cases. Thus, the differential diagnosis checklist, which is involved with the analytical process, displays its effectiveness for difficult or challenging cases rather than for easy cases. In contrast, there may be an advantage for intuitive process for simple cases in terms of its swiftness and efficiency. It is important to reappraise the importance of our intuition in order to diagnose common medical conditions.

In contrast, a checklist of differential diagnosis worked effectively for making accurate diagnosis for difficult or challenging cases. The one reason for this effectiveness might be that they included highlights on "Do not miss diagnosis" and "Frequently missed diagnosis". Although the checklist may work effectively, in reality, it seems difficult to use checklists for every single patient in hectic clinical environments, such as emergency rooms or walk-in clinics. The use of a differential diagnosis checklist may be better implemented for facing patients with difficult diagnosis.

In order to prove and generalize the validity of the usefulness for a differential diagnosis, further studies are needed to evaluate the usefulness of the checklists in actual clinical settings, and also to examine the effects by the use of checklists on clinical outcome for real patients. It may be needed to evaluate a number of cases that involve a broad spectrum of disease, ages, and ethnicities. Since our study involved only medical students, further studies are also needed to entail the involvement of physicians, including residents and staff physicians.

In conclusion, among medical students, the use of differential diagnosis checklists improved diagnostic performance in difficult cases but it might lead to reduced performance among simple cases. General de-biasing checklist did not help medical students for better diagnostic performance. Our results seem to be consistent with the dual processing theory implying that analytical process works better for difficult/challenging cases and intuitive process may be more effective for solving easy cases.

**References**


1) Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165(13):1493-1499.
2) Schiff GD, Hasan O, Kim S et al. Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors, Arch Intern Med. 2009;169(20):1881-1887.
3) Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primarycare—a systematic review. FamPract. 2008;25(6):400-413.
4) Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study 1. N Eng J Med. 1991;324:370–6.
5) Wilson RM, Runciman WB, Gibberd RW, et al. The Quality in Australian Health Care Study. Med J Australia 1995;163:458–71
6) Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000;38:261–2.
7) Tokuda Y, Kishida N, Konishi R, Koizumi S. Cognitive error as the most frequent

contributory factor in cases of medical injury: A study on verdict's judgment among closed claims in Japan. J Hosp Med. 2011; 6(3):109-14

8) Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med. 2006; 145(7):488-496.

9) Kahnemann D, Slovic P, Tversky A. Judgment Under Uncertainty: Heuristics and Biases. New York, NY: Cambridge University Press;1982:1-544

10) CroskerryP.The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78(8):775-780.

11) Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. J EvalClinPract. 2007;13(1):138-145.

12) Croskerry P. The cognitive imperative: thinking about how we think. AcadEmerg Med 2000;7:1223-31.

13) Croskerry P. Cognitive forcing strategies in clinical decision making. Ann Emerg Med 2003;41:110-9.

14) Evans JS. Dual-processing accounts of reasoning, judgment, and social cognition. Ann Rev Psychol. 2008;59:255-278.

15) Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Acad Med. 2008 Dec;83(12):1210-6.

16)Schiff GD. Minimizing diagnostic error: the importance of follow-up and feedback.Am J Med. 2008;121(5)(suppl):S38-S42.

17)Singh H, Thomas EJ, Khan MM, Petersen LA. Identifying diagnostic errors in pri primary care using an electronic screening algorithm. Arch Intern Med. 2007; 167(3):302-308.

18)Mamede. Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. JAMA 2010; 304(11):1198-1203

19)Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ. 2008;42(5):468-475.

20)Mamede S, Schmidt HG. The structure of reflective practice in medicine. Med Educ. 2004 Dec;38(12):1302-8.

21) Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. Med Educ. 2007 Dec;41(12):1185-92.

22) Mamede S, Schmidt HG. Correlates of reflective practice in medicine. Adv Health SciEduc Theory Pract. 2005 Nov;10(4):327-37.

23) Croskerry P. The feedback sanction. AcadEmerg Med. 2000;7:1232–38.

24) Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. Acad Med. 2011 Mar;86(3):307-13.

25) Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? Adv

Health SciEduc Theory Pract. 2009 Sep;14 Suppl 1:63-9. Epub 2009 Aug 11.

26) Stanovich KE. Who is Rational? Studies of Individual Differences in Reasoning. Mahwah, NJ: Lawrence Erlbaum 1999:1-312

27) Croskerry P.A universal model fordiagnostic reasoning. Acad Med. 2009; 84: 1022-28.

28) Medin D, Schaffer MM. A context theory of classification learning. Psychological Review 1978;85:207–38.

29) Norman G, Eva KW. Diagnostic error and clinical reasoning. Med Educ. 2010;44(1):94-100.

30) Ely JW, Graber ML, Croskerry PA. Checklist to reduce Diagnostic Errors. Acad Med 2011; 86:307-313

## Table 1

able 1: Score distribution of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC (N=18

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
|---|---|---|---|---|---|
| Group 1 (n=91) | | | | | |
| Intuitive diagnosis, n | | | | | |
| score 0 | 5 | 6 | 39 | 55 | 76 |
| 1 | 0 | 2 | 2 | 8 | 0 |
| 2 | 14 | 11 | 6 | 12 | 4 |
| 3 | 72 | 72 | 44 | 16 | 11 |
| Diagnosis by GDBC, n | | | | | |
| score 0 | 5 | 6 | 39 | 51 | 74 |
| 1 | 0 | 4 | 2 | 10 | 2 |
| 2 | 13 | 6 | 5 | 12 | 2 |
| 3 | 73 | 75 | 45 | 18 | 13 |
| Diagnosis by DDXC, n | | | | | |
| score 0 | 7 | 8 | 33 | 35 | 68 |
| 1 | 1 | 3 | 3 | 15 | 6 |
| 2 | 11 | 10 | 10 | 20 | 7 |
| 3 | 72 | 70 | 45 | 21 | 10 |
| Group 2 (n=97) | | | | | |
| Intuitive diagnosis, n | | | | | |
| score 0 | 11 | 11 | 49 | 70 | 90 |
| 1 | 3 | 5 | 3 | 3 | 2 |
| 2 | 3 | 8 | 9 | 8 | 3 |
| 3 | 80 | 73 | 36 | 16 | 2 |
| Diagnosis by DDXC, n | | | | | |
| score 0 | 12 | 20 | 39 | 58 | 71 |
| 1 | 3 | 2 | 5 | 9 | 6 |
| 2 | 6 | 7 | 5 | 11 | 4 |
| 3 | 76 | 68 | 48 | 19 | 16 |
| Diagnosis by GDBC, n | | | | | |
| score 0 | 12 | 17 | 38 | 56 | 71 |
| 1 | 2 | 5 | 7 | 9 | 4 |
| 2 | 7 | 4 | 4 | 11 | 5 |
| 3 | 76 | 71 | 48 | 21 | 17 |

Note: GDBC=general de-biasing checklist, DDXC=differential diagnosis checklist

## Table 2

Table 2: Mean scores of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC (N=188)

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Total cases |
|---|---|---|---|---|---|---|
| Group 1 (n=91) | | | | | | |
| Intuitive diagnosis | 2.68 | 2.64 | 1.60 | 0.88 | 0.45 | 8.25* |
| Diagnosis by GDBC | 2.69 | 2.65 | 1.62 | 0.97 | 0.49 | 8.42 |
| Diagnosis by DDXC | 2.63 | 2.56 | 1.74 | 1.30 | 0.55 | 8.77* |
| Group 2 (n=97) | | | | | | |
| Intuitive diagnosis | 2.57 | 2.47 | 1.33 | 0.69 | 0.14 | 7.21** |
| Diagnosis by DDXC | 2.51 | 2.27 | 1.64 | 0.91 | 0.64 | 7.96** |
| Diagnosis by GDBC | 2.52 | 2.33 | 1.64 | 0.97 | 0.67 | 8.12 |

*P=0.01 by Sidak's analysis based on repeated measures of ANOVA adjusted for gender and grade.

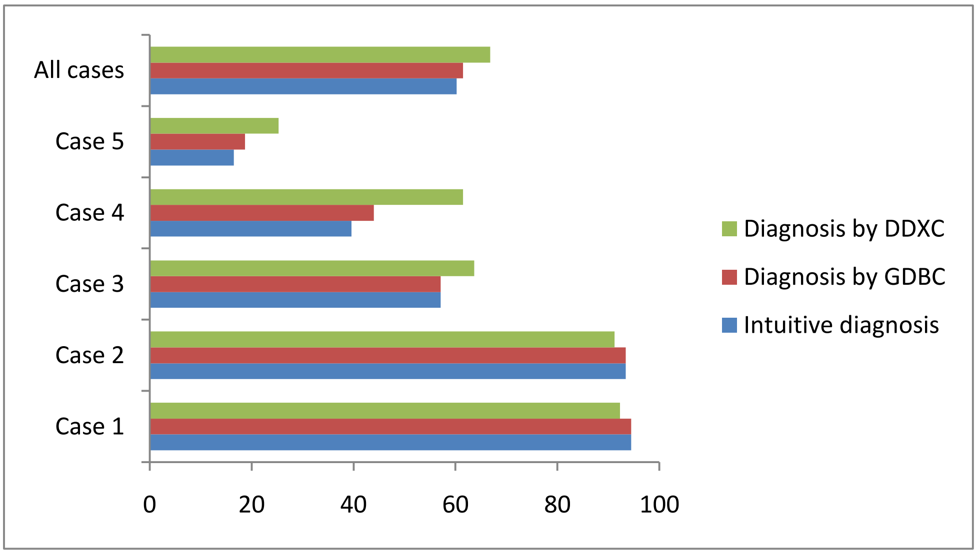
**P=0.001 by Sidak's analysis based on repeated measures of ANOVA adjusted for gender and grade.

Note: GDBC=general de-biasing checklist, DDXC=differential diagnosis checklist

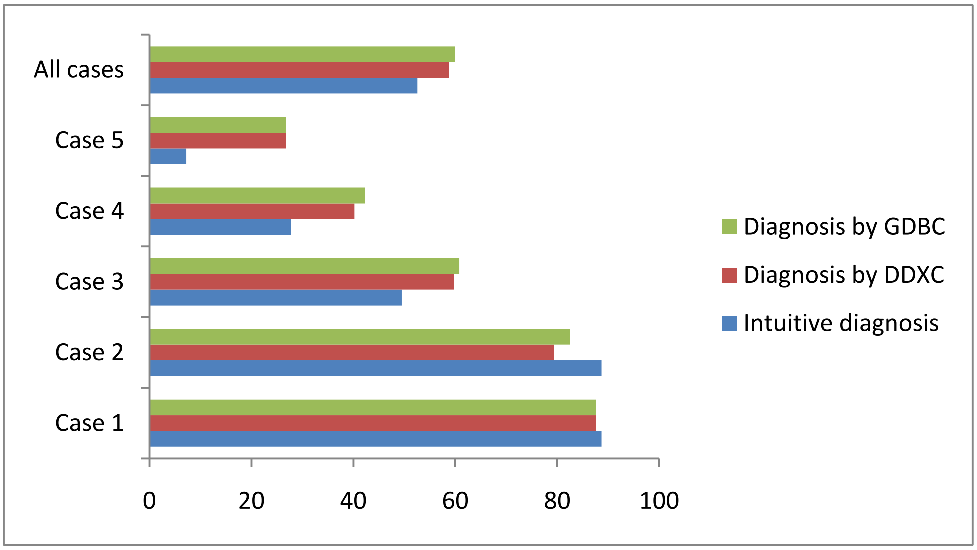
**Figure 1**

**Title:** Proportions of correct diagnosis of intuitive diagnosis, diagnosis by GDBC, and diagnosis by DDXC by the groups 1 and 2
(Horizontal axis indicates the proportions of correct diagnosis)

**Group 1**




**Group 2**




Note: GDBC=general de-biasing checklist, DDXC=differential diagnosis checklist