

Original Paper

# Incidence of Diagnostic Errors Among Unexpectedly Hospitalized Patients Using an Automated Medical History–Taking System With a Differential Diagnosis Generator: Retrospective Observational Study

Ren Kawamura<sup>1</sup>, MD; Yukinori Harada<sup>1,2</sup>, MD, PhD; Shu Sugimoto<sup>2</sup>, MD; Yuichiro Nagase<sup>2</sup>, MD; Shinichi Katsukura<sup>1</sup>, MD; Taro Shimizu<sup>1</sup>, MPH, MBA, MD, PhD

<sup>1</sup>Department of Diagnostic and Generalist Medicine, Dokkyo Medical University, Mibu, Japan

<sup>2</sup>Department of Internal Medicine, Nagano Chuo Hospital, Nagano, Japan

**Corresponding Author:**

Taro Shimizu, MPH, MBA, MD, PhD

Department of Diagnostic and Generalist Medicine

Dokkyo Medical University

880 Kitakobayashi

Mibu, 321-0293

Japan

Phone: 81 282861111

Email: [shimizutaro7@gmail.com](mailto:shimizutaro7@gmail.com)

## Abstract

**Background:** Automated medical history–taking systems that generate differential diagnosis lists have been suggested to contribute to improved diagnostic accuracy. However, the effect of these systems on diagnostic errors in clinical practice remains unknown.

**Objective:** This study aimed to assess the incidence of diagnostic errors in an outpatient department, where an artificial intelligence (AI)–driven automated medical history–taking system that generates differential diagnosis lists was implemented in clinical practice.

**Methods:** We conducted a retrospective observational study using data from a community hospital in Japan. We included patients aged 20 years and older who used an AI-driven, automated medical history–taking system that generates differential diagnosis lists in the outpatient department of internal medicine for whom the index visit was between July 1, 2019, and June 30, 2020, followed by unplanned hospitalization within 14 days. The primary endpoint was the incidence of diagnostic errors, which were detected using the Revised Safer Dx Instrument by at least two independent reviewers. To evaluate the effect of differential diagnosis lists from the AI system on the incidence of diagnostic errors, we compared the incidence of these errors between a group where the AI system generated the final diagnosis in the differential diagnosis list and a group where the AI system did not generate the final diagnosis in the list; the Fisher exact test was used for comparison between these groups. For cases with confirmed diagnostic errors, further review was conducted to identify the contributing factors of these errors via discussion among three reviewers, using the Safer Dx Process Breakdown Supplement as a reference.

**Results:** A total of 146 patients were analyzed. A final diagnosis was confirmed for 138 patients and was observed in the differential diagnosis list from the AI system for 69 patients. Diagnostic errors occurred in 16 out of 146 patients (11.0%, 95% CI 6.4%–17.2%). Although statistically insignificant, the incidence of diagnostic errors was lower in cases where the final diagnosis was included in the differential diagnosis list from the AI system than in cases where the final diagnosis was not included in the list (7.2% vs 15.9%,  $P=.18$ ).

**Conclusions:** The incidence of diagnostic errors among patients in the outpatient department of internal medicine who used an automated medical history–taking system that generates differential diagnosis lists seemed to be lower than the previously reported incidence of diagnostic errors. This result suggests that the implementation of an automated medical history–taking system that generates differential diagnosis lists could be beneficial for diagnostic safety in the outpatient department of internal medicine.

(*JMIR Med Inform* 2022;10(1):e35225) doi: [10.2196/35225](https://doi.org/10.2196/35225)

**KEYWORDS**

artificial intelligence; automated medical history-taking; diagnostic errors; outpatient; Safer Dx

## Introduction

Diagnostic error, defined as the failure to establish an accurate and timely explanation of the patient's health problem or to communicate that explanation to the patient [1], is one of the most important patient safety issues that should be addressed [2,3]. The impact of diagnostic errors on patient safety is quite large [4]. First, diagnostic errors comprise around 50% of preventable harm in primary health care settings and emergency departments [5]. Second, the risk of death, significant permanent injury, and prolonged hospitalization is higher for diagnostic error cases than for other medical errors [6-12]. Third, diagnostic errors frequently occur in several settings of clinical practice; approximately 5% of patients can experience diagnostic errors in primary health care and hospital practice in the United States [13]. Therefore, effective interventions to reduce diagnostic errors are warranted.

Diagnostic error-related paid malpractice claims occur more often among outpatients than among inpatients [9], suggesting that the primary health care outpatient setting is vulnerable to diagnostic errors. The prevalence of diagnostic errors in outpatient settings has been reported to be between 3.6% and 5.1%. However, when focusing on a population of patients with a high risk for diagnostic errors who were unexpectedly hospitalized within 14 days after the index outpatient visit, the prevalence of diagnostic errors increased to as much as 21% [14]. The common contributing factors for diagnostic errors in primary care outpatient settings were reported to include problems with history-taking, overreliance on pattern recognition, and failure to consider sufficient differential diagnoses [4,15]. Therefore, strategies or systems to improve the quality of history-taking and support differential diagnosis generation are required to reduce diagnostic errors in outpatient settings.

From this perspective, newly developed technology, such as computerized automated history-taking systems and diagnostic decision support systems, can be leveraged to address this issue; these systems have a long history, since they were introduced in the 1960s and 1970s [16-18]. Computerized automated history-taking systems perform better in clinical documentation tasks for taking patient histories than do physicians [19,20]. The use of a diagnostic support system (ie, differential diagnosis generator) before collecting information by physicians showed a significant impact on the improvement of diagnostic accuracy in terms of clinical reasoning and differential diagnosis [21-23]. Moreover, a new system that combines automated medical history-taking functions with differential diagnosis generation—specialized for musculoskeletal diseases only—showed improved diagnostic accuracy among physicians in a pilot randomized controlled trial [24]. Subsequently, another system, covering broad symptoms of internal diseases, was developed and implemented in clinical practice [25]. Yet another study showed high reliability of documentation regarding clinical history to assist the diagnostic accuracy of physicians

[26]; however, this was not conducted in a clinical practice setting.

These automated systems have generated concerns about their negative effects on the diagnostic accuracy of physicians. For instance, physicians may not accept correct diagnoses or may accept incorrect diagnoses generated by the systems [24,26], partly because physicians tend to be more confident with their own diagnosis than that of artificial intelligence (AI) systems when there is a discrepancy between them [27]. Therefore, the effects of the implementation of these systems on diagnostic errors in clinical practice remain unknown. This study aimed to assess the incidence of diagnostic errors in an outpatient department, where an AI-driven automated medical history-taking system that generates differential diagnosis lists was implemented in clinical practice.

## Methods

### Study Design

We conducted a retrospective observational study using data from Nagano Chuo Hospital in Japan. The Research Ethics Committee of Nagano Chuo Hospital approved this study (serial number: NCR202104). The requirement to obtain written informed consent from patients was waived by the Research Ethics Committee under the condition that we used an opt-out method. We informed patients by showing the detailed information of the study on the official website of Nagano Chuo Hospital.

### Patient Population

We included patients aged 20 years and older who used AI Monshin—an AI-based automated medical history-taking system—in the outpatient department of internal medicine for whom the index visit was between July 1, 2019, and June 30, 2020, followed by unplanned hospitalization within 14 days. A follow-up duration of 14 days was selected to improve the sensitivity to detect diagnostic errors [14,28]. For assessing the effects of using AI Monshin on diagnostic errors, we excluded patients for whom AI Monshin did not list 10 differential diagnoses. In those cases, the AI system could not complete history-taking because patients gave up entering information or because they presented to the hospital for further investigation of abnormal test results following their annual health checkup, which was out of scope for the system during the study period. Usually, even one differential diagnosis was not generated in such cases.

### Presentation of the AI Monshin Tool

The details of AI Monshin were presented in a previous report [25]. In brief, AI Monshin converts data entered by patients on tablet terminals into medical terms. Patients enter their background information, such as age and sex, and chief complaint as free text on a tablet in the waiting room. AI Monshin asks approximately 20 questions, one by one, which are tailored to the patient. The questions are optimized, based on previous answers, to generate the most relevant list of

potential differential diagnoses. Physicians can see the entered data as a summarized medical history with the top 10 possible differential diagnoses, along with their rank.

### Identification of Diagnostic Errors

To identify whether diagnostic errors occurred in this study, we used the Revised Safer Dx Instrument [29]. The Safer Dx Instrument is an externally validated, structured data collection tool to improve the accuracy of assessment of diagnostic errors [30,31]; the tool has been widely used in several studies on diagnostic errors [32-36]. Recently, the tool was updated as the Revised Safer Dx Instrument [29]. The Revised Safer Dx Instrument consists of 13 items. Items 1 to 12 are used for assessing the diagnostic process, and item 13 is used to determine the possibility of diagnostic error. All items are rated by answering questions on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). The Revised Safer Dx Instrument can be used to assess the entire diagnostic process of one event; however, because we focused on diagnostic errors related to the implementation of AI Monshin, which seems to mainly influence the diagnostic decision at the index visit, the evaluation of diagnostic errors in this study was based on the medical records taken during the index visit.

The identification of diagnostic errors in this study was conducted through the algorithm as discussed in this section. In the first step, two reviewers (YH and SS) independently evaluated the diagnostic process of included cases using the Revised Safer Dx Instrument by reviewing the medical records. The presence or absence of diagnostic errors in each case was judged based on the score of item 13 [29]. According to the recommendation for using the Revised Safer Dx Instrument, diagnostic error was confirmed in cases where both reviewers scored 5 or higher on item 13, and diagnostic error was denied in cases where both reviewers scored 3 or lower on item 13 [29]. The remaining cases were progressed to the second step. In the second step, the third reviewer (YN) independently evaluated the cases using the Revised Safer Dx Instrument. Diagnostic error was confirmed in cases where two out of three reviewers scored 5 or higher on item 13, and diagnostic error was denied in cases where two out of three reviewers scored 3 or lower on item 13. For the remaining cases in which diagnostic error was neither confirmed nor denied, the three reviewers (YH, SS, and YN) discussed and mutually agreed on whether diagnostic error occurred or not on a case-by-case basis.

The final diagnoses of all cases were confirmed by two reviewers (YH and SS) based on the discharge summary. Disagreements were resolved by discussion among the three reviewers (YH, SS, and YN). Based on the confirmed final diagnoses, the other two reviewers (RK and SK), who were blinded to the evaluation of diagnostic errors, independently judged whether the final diagnosis of each case was included in the list of 10 differential diagnoses generated by AI Monshin. Disagreements were resolved by discussion between the two reviewers (RK and SK).

### Analysis of the Causes of Diagnostic Errors

For cases with confirmed diagnostic errors, further review was conducted to identify the contributing factors of these errors

via discussion among the three reviewers (YH, SS, and YN). The Safer Dx Process Breakdown Supplement was used as a reference to classify the contributing factors of diagnostic errors and outcomes in this study [29]. To evaluate the effects of AI Monshin implementation on the diagnostic errors, other than the items in the Safer Dx Process Breakdown Supplement, the following were discussed: the frequency of the final diagnosis (ie, whether the disease was common or uncommon), typicality of the presentation for the final diagnosis (ie, typical or atypical), and initial diagnosis at the index visit.

### Baseline Data Collection and Outcome

From the medical records, we extracted data on the age and sex of patients, chief complaints, and the experience of physicians who saw patients at the index visits (ie, resident: up to 5 years of experience after graduation; staff: more than 5 years of experience after graduation). The primary outcome was the incidence of diagnostic errors.

### Sample Size Calculation

We calculated the required sample size to be 139 cases, with an incidence of diagnostic errors of 10.0% and a margin of 5.0%. It was estimated that there were approximately 150 patients who were eligible for this study between July 1, 2019, and June 30, 2020. Even with the expectation that approximately 5 to 10 cases could be excluded, 150 cases were a reasonable target number of cases for this study.

### Statistical Analysis

Continuous data are presented as medians with the 25th and 75th percentiles. Categorical data are presented as counts and proportions (%). For the primary outcome, we calculated the incidence of diagnostic errors with 95% CI. To evaluate the baseline factors and the differential diagnosis list of AI Monshin with regard to the incidence of diagnostic errors, we compared the incidence of diagnostic errors between the groups of older adults (aged  $\geq 65$  years) and non-older adults (aged  $< 65$  years) [37-40], the groups of males and females [33], the groups seen by staff and seen by residents [26], and the groups in which AI Monshin generated or did not generate the final diagnosis in the differential diagnosis list [26]; these comparisons were made using the Fisher exact test. We also calculated the odds ratio (OR) with 95% CI for the incidence of diagnostic errors in these groups. *P* values were based on 2-tailed statistical tests, and *P* values less than .05 were considered statistically significant. All statistical analyses were conducted using R (version 4.1.0; The R Foundation).

## Results

### Baseline Patient Characteristics

A total of 150 cases were unexpectedly hospitalized within 14 days after the index visit that took place at the outpatient department of internal medicine; AI Monshin was used at the index visit. Only 2 (1.3%) patients did not complete history-taking by AI Monshin: a woman in her 70s complained of an uncomfortable feeling on her tongue, abdominal pain with distention, and appetite loss, and a man in his 70s complained that his cold was not getting better. After excluding 4 (2.7%)

cases in which AI Monshin did not develop 10 differential diagnoses (2 cases: incomplete history-taking; 2 cases: patients presented for further investigation for abnormal test results), the data from 146 cases were analyzed for this study. The median age of the patients was 71 (IQR 59-82) years, 72 (49.3%) were male, 71 (48.6%) were seen by residents at the index visit, and 103 (70.5%) were admitted to the hospital on the same day as the index visit.

**Chief Complaints and the Final Diagnosis**

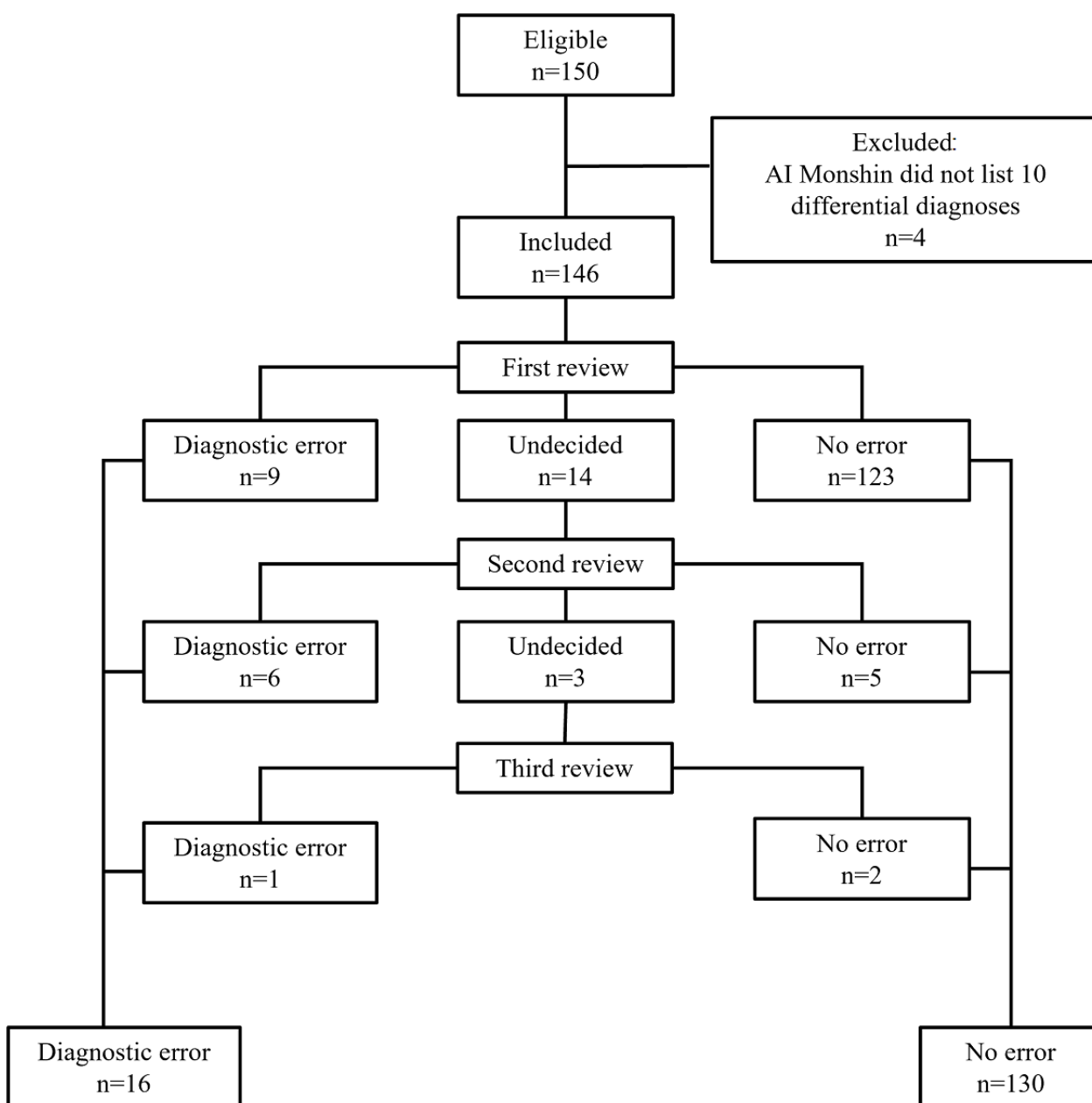
The top three most common chief complaints were abdominal pain (37/146, 25.3%), fever (20/146, 13.7%), and melena or hematochezia (15/146, 10.3%). During follow-up outpatient visits or admission, the final diagnosis was confirmed for 138 patients (94.5%). The most common diagnosis was lower respiratory tract infection (15/138, 10.9%), followed by ischemic

colitis (8/138, 5.8%), diverticular bleeding (8/138, 5.8%), and congestive heart failure (8/138, 5.8%). The final diagnosis was based on the differential diagnosis list from AI Monshin for 69 out of 138 patients (50.0%).

**Primary Outcome**

Figure 1 shows the steps of the review for confirming the diagnostic errors in this study. In the first step of the review, diagnostic errors were confirmed in 9 cases and denied in 123 cases. Among the remaining 14 cases, diagnostic errors were confirmed in 6 cases and denied in 5 cases in the second step of the review. Among the remaining 3 cases, diagnostic errors were confirmed in 1 case and denied in 2 cases in the third step of the review. In total, diagnostic errors were confirmed in 16 out of 146 cases (11.0%, 95% CI 6.4%-17.2%).

**Figure 1.** Flow of reviews for confirming diagnostic errors. AI: artificial intelligence.



The incidence of diagnostic errors was significantly higher in patients aged 65 years and older compared to those under 65 years of age (15/96, 16% vs 1/50, 2%; OR 9.1, 95% CI 1.2-70.8;  $P=.01$ ). There were no significant differences in the incidence of diagnostic errors between male and female patients (11/72, 15% vs 5/74, 7%; OR 2.5, 95% CI 0.8-7.6;  $P=.12$ ), between patients who were seen by a resident and those who were seen by a physician at the index visit (9/71, 13% vs 7/75, 9%; OR 1.4, 95% CI 0.5-4.0;  $P=.60$ ), and between cases in which the final diagnosis was not included in the differential diagnosis list from AI Monshin and those in which the final diagnosis was included in the same list (11/69, 16% vs 5/69, 7%; OR 2.4, 95% CI 0.8-7.4;  $P=.18$ ).

### Details Regarding Cases With Diagnostic Errors

Table 1 and Multimedia Appendix 1 show the details of the 16 cases where there were diagnostic errors. All cases had common final diagnoses (ie, cholangitis, cholecystitis, diverticular bleeding, pneumonia, interstitial pneumonia, intestinal obstruction, pyelonephritis, infectious enteritis, heart failure, and pulmonary artery embolism), and the final diagnosis presentation was typical for 15 out of 16 cases (94%). The most common chief complaint in the 16 cases with diagnostic errors was abdominal pain ( $n=5$ , 31%), followed by cough ( $n=4$ , 25%) and fever ( $n=3$ , 19%).

According to the Safer Dx Process Breakdown Supplement, the most common contributing factors for diagnostic errors in 16 cases were “problems ordering diagnostic tests for further workup” ( $n=13$ , 81%), followed by “problems with data

integration and interpretation” ( $n=10$ , 63%), “problems with physical exam” ( $n=9$ , 56%), and “performed tests not interpreted correctly” ( $n=8$ , 50%; Table 2).

From the aspect of the differential diagnosis list for cases with diagnostic errors, AI Monshin listed the final diagnosis in the list in 5 out of 16 cases (31%) and the initial diagnosis in 4 out of 16 cases (25%). On the other hand, in cases without diagnostic errors, AI Monshin listed the final diagnosis in the differential list in 64 out of 122 cases (52.5%, excluding 8 cases where the final diagnosis was unknown). In summary, despite using AI Monshin, physicians could not make the correct diagnoses as were suggested in the differential diagnosis list in 5 of 69 cases (7% omission errors). On the other hand, the incorrect initial diagnoses made by physicians were listed in the differential diagnosis list in 4 of 69 cases (6% commission errors). Regarding the outcome, no cases of diagnostic errors resulted in death or permanent harm. A total of 2 cases out of 16 (13%) were classified as Category C: “An error occurred that reached the patient but did not cause the patient harm.” Diagnostic errors resulted in some harm in 14 out of 16 cases (88%; 2 cases were classified as Category E: “An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention”; 12 cases were classified as Category F: “An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization”). The median time between the index visit and the time that the final diagnosis was made was 3 (IQR 2-6) days.

**Table 1.** The details of 16 diagnostic error cases.

Case No. <sup>a</sup>	Age (y)	Sex <sup>b</sup>	Physician of first contact	Chief complaint	Initial diagnosis	Final diagnosis	Index visit to final diagnosis (days), n	Outcome category <sup>c</sup>	Initial diagnosis was on list <sup>d</sup>	Final diagnosis was on list <sup>d</sup>
1	95	F	Resident	Fever	URI <sup>e</sup>	Cholangitis	4	F	No	No
2	76	M	Resident	Abdominal pain	GERD <sup>f</sup>	Cholecystitis	2	F	Yes; rank 4	No
3	83	M	Resident	Abdominal pain	Costo-chondritis	Pneumonia	3	F	No	No
4	55	M	Resident	Hematochezia	Infectious enteritis	Diverticular bleeding	2	F	Yes; rank 3	Yes; rank 1
5	89	F	Staff	Nausea	Unknown	Acute pyelonephritis	3	F	No	No
6	75	M	Staff	Cough	URI	Interstitial pneumonia	3	F	No	Yes; rank 10
7	66	M	Resident	Abdominal pain	Constipation	Intestinal obstruction	6	F	Yes; rank 4	No
8	70	F	Staff	Cough	Unknown	Heart failure	3	F	No	Yes; rank 8
9	77	F	Resident	Palpitation	Heart failure	Pulmonary embolism	2	E	Yes; rank 10	No
10	82	M	Staff	Fever	URI	Cholecystitis	3	F	No	No
11	81	F	Resident	Anorexia	Cholelithiasis	Acute pyelonephritis	2	C	No	No
12	72	M	Staff	Headache, lightheadedness	Fatigue	Vestibular neuritis	8	E	No	No
13	86	M	Resident	Abdominal pain	Enteritis	Intestinal obstruction	0 <sup>g</sup>	F	No	Yes; rank 9
14	78	M	Staff	Abdominal pain	Hemorrhoid	Infectious enteritis	9	C	No	No
15	91	M	Staff	Fever, cough, back pain	URI	Acute pyelonephritis	7	F	No	Yes; rank 3
16	72	M	Resident	Dyspnea, cough, malaise	URI	Interstitial pneumonia	11	F	No	No

<sup>a</sup>All diagnoses were common. All cases had typical presentations except for case 2.

<sup>b</sup>Female (F) or male (M).

<sup>c</sup>Outcome was classified, along with the Safer Dx Process Breakdown Supplement, as follows: Category C, "An error occurred that reached the patient but did not cause the patient harm"; Category E, "An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention"; Category F, "An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization" [29].

<sup>d</sup>AI Monshin's differential list; where a diagnosis was on the list, its rank on the list is indicated.

<sup>e</sup>URI: upper respiratory infection.

<sup>f</sup>GERD: gastroesophageal reflux disease.

<sup>g</sup>The final diagnosis was made at the second visit, which was on the same day as the index visit.

**Table 2.** Breakdown analysis of the contributing factors for diagnostic errors.

Contributing factors and details	Cases (N=16), n (%)
<b>Patient-related factors</b>	
Delay in seeking care	0 (0)
Lack of adherence to appointments	0 (0)
Other	0 (0)
<b>Patient-provider encounter</b>	
Problems with history	4 (25)
Problems with physical exam	9 (56)
Problems ordering diagnostic tests for further workup	13 (81)
Failure to review previous documentation	4 (25)
Problems with data integration and interpretation	10 (63)
Other	0 (0)
<b>Diagnostic tests</b>	
Ordered test was not performed at all	0 (0)
Ordered test was not performed correctly	0 (0)
Performed test was not interpreted correctly	8 (50)
Misidentification	1 (6)
Other	0 (0)
<b>Follow-Up and tracking</b>	
Problems with timely follow-up of abnormal diagnostic test results	1 (6)
Problems with scheduling of appropriate and timely follow-up visits	2 (13)
Problems with diagnostic specialties returning test results to clinicians	2 (13)
Problems with clinicians reviewing test results	0 (0)
Problems with clinicians documenting action or response to test results	0 (0)
Problems with notifying patients of test results	0 (0)
Problems with monitoring patients through follow-up	0 (0)
Other	0 (0)
<b>Referrals</b>	
Problems initiating referral	1 (6)
Lack of appropriate actions on requested consultation	0 (0)
Communication breakdown from consultant to referring provider	0 (0)
Other	0 (0)

## Discussion

### Principal Findings

Among 146 patients who used the AI-driven, automated history-taking system, which developed a list of the top 10 differential diagnoses, diagnostic errors occurred in 11.0% of cases. These patient histories were collected at the index visit to the outpatient department of internal medicine, followed by unplanned hospitalization of the patient within 14 days. The incidence of diagnostic errors was statistically higher among older adult patients; however, the sex of the patients, the experience of the physicians, and the accuracy of the differential diagnosis list of the AI system were not statistically associated

with the incidence of diagnostic errors. In all cases where diagnostic errors occurred, the final diagnoses were common diseases, as reported in a previous study that was conducted in primary care settings in the United States between 2006 and 2007 [4], and the clinical presentation was typical, except in one case.

### Limitations

To the best of our knowledge, this is the first observational study that evaluated the effects of implementation of an automated medical history-taking system with a differential diagnosis generator in routine clinical practice using the validated Revised Safer Dx Instrument to detect diagnostic errors. However, this study also had some limitations. First, this

study did not include patients who did not use an automated history-taking system with a differential diagnosis generator or those who were not admitted; therefore, the incidence of diagnostic errors should be interpreted with caution. Second, exclusion of the cases in which AI Monshin did not develop 10 differential diagnoses may have reduced the incidence of diagnostic errors in this study. Since inadequate and inappropriate history could be a contributing factor for diagnostic errors, excluding such a case may merit the optimistic assumption of AI Monshin's performance. Third, because the judgment of diagnostic errors was conducted by a retrospective review of the charts, some bias could not be avoided. However, as the review process was predefined and at least two reviewers independently assessed each case, we are sure that these biases were avoided as much as possible. Fourth, we are unsure of the effects of COVID-19 on diagnostic errors in the outpatient department. Future studies may focus on the incidence of diagnostic errors between hospitals with and without implementation of an automated medical history-taking system with a diagnostic decision support function in a prospective design.

### Comparison With Prior Work

The incidence of diagnostic errors in this study was 11.0%, which was lower than that reported in previous studies (13.7% and 20.9%) that included cases similar to this study (ie, patients who were unexpectedly hospitalized within 14 days after their index visit) [14,28]. In addition, the incidence of diagnostic errors in this study was lower than that reported in retrospective studies with chart review (13.3% to 21.8%) [11,41-43] or in prospective studies (12.3% to 20.0%) [12,44] that investigated the rate of discrepancy in the diagnosis between admission and discharge. Therefore, it is possible that the implementation of an automated history-taking system with a differential diagnosis generator reduced the incidence of diagnostic errors in the outpatient department of internal medicine.

The quality of clinical history documented by AI Monshin may be a key component of the results. There may be high discrepancies in clinical history between patient reports and physician documentation [45]; in addition, the automated medical history-taking system, as compared to physicians, may have the potential to take clinical histories that are more diagnostically useful and of higher quality [19,20]. Therefore, routine use of automated history-taking systems may improve diagnostic accuracy by establishing a high-quality base of clinical history for the correct diagnosis. Indeed, in a previous study that used the documentation made by an automated medical history-taking system from real patients, the correct diagnosis appeared in 56.3% of the top three differential diagnoses made by physicians without using a differential diagnosis list from an AI-driven system; this increased to 72.7% in cases where the correct diagnosis was included in the AI-driven differential diagnosis list [26]. Furthermore, a previous study of another automated medical history-taking system with a differential diagnosis generator—DIAANA, specializing in injury or disease of the musculoskeletal system—showed that the diagnostic accuracy was superior in the group in which physicians used the system compared to the group in which physicians did not use the system; this was a

pilot randomized controlled trial conducted in a real clinical practice setting [24]. In contrast to the previous study that identified history-taking as the most common contributing factor of diagnostic errors [4], the breakdown analysis of the diagnostic errors in this study did not identify history-taking as the main contributing factor of these errors, indicating that the implementation of an automated history-taking system with diagnostic decision support could reduce the diagnostic errors associated with poor clinical history-taking.

In addition to making a high-quality document of medical history, an automated medical history-taking system with a differential diagnosis generator seems to have some advantages. First, this system can be integrated into routine diagnostic processes in clinical practice. Currently, one of the most important concerns in the diagnostic decision support system is its low usage rate. For example, in the case of Isabel, which is one of the most famous AI-driven diagnostic decision support systems that generates a differential diagnosis list based on entered information by physicians, a previous study showed that only 7.9% of participants who were given open access to Isabel reported using Isabel at least once a week, whereas the others never used it [46]. According to the other two studies, on average, Isabel was used for only 3 out of 4840 patients (0.06%) for 3 months [47], and the usage rate did not increase despite frequent reminders for clinicians to use Isabel on a regular basis [48]. Such low use of a diagnostic decision support system appeared to be caused by physicians who did not recognize the need for diagnostic support, relying on their own acumen to deliver the correct diagnosis [49]. However, diagnostic decision support systems should operate seamlessly in the background in the diagnostic process in clinical practice, regardless of whether the physicians need it or not [49]. An automated medical history-taking system with a differential diagnosis generator can address such an unmet need and may reduce diagnostic errors through routine support. Second, the use of a diagnostic decision support system at the early stage of the diagnostic process was reported to be more useful than its use at a later stage. To date, several studies have been conducted to evaluate the impact of the timing of using a diagnostic decision support system. According to their studies, physician diagnosis was associated with their first impression [50], and early use of diagnostic support systems before collecting information by physicians significantly improved the diagnostic accuracy [21-23]. These findings may support the positive effects of the implementation of an automated medical history-taking system with a differential diagnosis generator, which can provide diagnostic decision support before physicians collect information. Third, an automated medical history-taking system with a differential diagnosis generator can be used without additional time consumption. Another barrier for clinicians to use diagnostic decision support systems in routine clinical practice is time constraint, as previous studies have shown that using Isabel usually requires an additional 4 to 7 minutes per case [47,48]. On the other hand, an automated history-taking system with a differential diagnosis generator increased only 0.3 minutes of examination time per case in an internal medicine outpatient department [25]. Therefore, clinicians can use automated history-taking systems with



differential diagnosis generators without wasting additional time.

Furthermore, several limitations exist regarding the implementation of automated history-taking systems with differential diagnosis generators. First, at present, the accuracy of differential diagnosis lists of AI systems is not sufficiently high to believe the lists every time. A previous study reported that the prevalence of the correct diagnosis in the top 10 list of differential diagnoses from diagnostic decision support systems in clinical practice settings was around 50% [51]; similar to that study, the correct diagnosis appeared in only 50% of the top 10 lists of differential diagnoses from AI Monshin in this study. As an a priori incorrect diagnosis before a patient encounter can lead physicians to an incorrect final diagnosis [52], the relatively low accuracy of the differential diagnosis list from AI Monshin may prevent the positive effect of the implementation of an automated history-taking system with a differential diagnosis generator on the reduction of diagnostic errors. Although statistically insignificant, the incidence of diagnostic errors in cases where the correct diagnosis was included in the differential diagnosis list from the AI system was twice as high as that in cases where the correct diagnosis was not included in the list. However, among the 69 cases in which the final diagnosis was not included in the differential diagnosis list from the AI system, an incorrect diagnosis by a physician was observed in the differential diagnosis list from AI Monshin in only 4 cases (6%). In addition, a previous study showed that only 15% of physicians' diagnoses seemed to be associated with the differential diagnosis list from the AI system [53]. This indicates that the majority of diagnostic errors in this study were not

related to the incorrect differential diagnosis list from the AI system. Second, the correct diagnosis in the automated differential diagnosis list cannot always be accepted as the most likely diagnosis by a physician. In 5 out of 69 cases (7%) where the correct diagnosis was included in the AI-generated differential diagnosis list, the correct diagnosis was not accepted as the initial diagnosis by the physician in this study. However, this type of error was also lower than that reported in previous studies (10.0% and 15.9%) [24,53]. Third, automated medical history-taking systems have had difficulty in precise history-taking for specific patients, such as older adult patients [54]. Indeed, in cases with diagnostic errors in this study, important past medical history was not imputed for 3 patients. However, such missed information seemed to be easily covered by physicians by checking the past medical history directly from the patient or reviewing the previous documentation.

## Conclusions

The incidence of diagnostic errors seems to be reduced by the implementation of an automated medical history-taking system with a diagnostic decision support function in the outpatient department. Although the accuracy of the differential diagnosis list from AI Monshin remains low, the negative effects of incorrect differential diagnosis lists from AI systems on the diagnostic accuracy of physicians could be counteracted by the high-quality clinical history taken by AI systems. Therefore, in total, the implementation of an automated history-taking system with diagnostic decision support may have more beneficial impacts than negative effects on diagnostic safety in the outpatient department.

## Acknowledgments

This work was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (KAKENHI) program (grant JP21K10355).

## Authors' Contributions

RK and YH were responsible for conceptualization of the study and for developing the study methodology. YH conducted the formal analysis, was responsible for securing resources, performed data curation, and was responsible for project administration and funding acquisition. RK, YH, SS, YN, and SK conducted the study investigation. RK was responsible for writing and preparing the original draft of the manuscript. YH and TS were responsible for reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Details of the histories written by AI Monshin in 16 diagnostic error cases. AI: artificial intelligence.

[\[DOCX File , 32 KB-Multimedia Appendix 1\]](#)

## References

1. Balogh EP, Miller BT, Ball JR, editors. Overview of diagnostic error in health care. In: Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015:81-144.
2. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf* 2017 Jun;26(6):484-494 [FREE Full text] [doi: [10.1136/bmjqs-2016-005401](https://doi.org/10.1136/bmjqs-2016-005401)] [Medline: [27530239](https://pubmed.ncbi.nlm.nih.gov/27530239/)]
3. Cresswell KM, Panesar SS, Salvilla SA, Carson-Stevens A, Larizgoitia I, Donaldson LJ, World Health Organization's (WHO) Safer Primary Care Expert Working Group. Global research priorities to better understand the burden of iatrogenic

- harm in primary care: An international Delphi exercise. *PLoS Med* 2013 Nov;10(11):e1001554 [FREE Full text] [doi: [10.1371/journal.pmed.1001554](https://doi.org/10.1371/journal.pmed.1001554)] [Medline: [24260028](https://pubmed.ncbi.nlm.nih.gov/24260028/)]
4. Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013 Mar 25;173(6):418-425 [FREE Full text] [doi: [10.1001/jamainternmed.2013.2777](https://doi.org/10.1001/jamainternmed.2013.2777)] [Medline: [23440149](https://pubmed.ncbi.nlm.nih.gov/23440149/)]
  5. Fernholm R, Pukk Härenstam K, Wachtler C, Nilsson GH, Holzmann MJ, Carlsson AC. Diagnostic errors reported in primary healthcare and emergency departments: A retrospective and descriptive cohort study of 4830 reported cases of preventable harm in Sweden. *Eur J Gen Pract* 2019 Jul;25(3):128-135 [FREE Full text] [doi: [10.1080/13814788.2019.1625886](https://doi.org/10.1080/13814788.2019.1625886)] [Medline: [31257959](https://pubmed.ncbi.nlm.nih.gov/31257959/)]
  6. Gupta A, Snyder A, Kachalia A, Flanders S, Saint S, Chopra V. Malpractice claims related to diagnostic errors in the hospital. *BMJ Qual Saf* 2017 Aug 09;27(1):53-60. [doi: [10.1136/bmjqs-2017-006774](https://doi.org/10.1136/bmjqs-2017-006774)] [Medline: [28794243](https://pubmed.ncbi.nlm.nih.gov/28794243/)]
  7. Watari T. Malpractice claims of internal medicine involving diagnostic and system errors in Japan. *Intern Med* 2021 Sep 15;60(18):2919-2925 [FREE Full text] [doi: [10.2169/internalmedicine.6652-20](https://doi.org/10.2169/internalmedicine.6652-20)] [Medline: [33776001](https://pubmed.ncbi.nlm.nih.gov/33776001/)]
  8. Watari T, Tokuda Y, Mitsuhashi S, Otuki K, Kono K, Nagai N, et al. Factors and impact of physicians' diagnostic errors in malpractice claims in Japan. *PLoS One* 2020;15(8):e0237145 [FREE Full text] [doi: [10.1371/journal.pone.0237145](https://doi.org/10.1371/journal.pone.0237145)] [Medline: [32745150](https://pubmed.ncbi.nlm.nih.gov/32745150/)]
  9. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: An analysis from the National Practitioner Data Bank. *BMJ Qual Saf* 2013 Aug;22(8):672-680. [doi: [10.1136/bmjqs-2012-001550](https://doi.org/10.1136/bmjqs-2012-001550)] [Medline: [23610443](https://pubmed.ncbi.nlm.nih.gov/23610443/)]
  10. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Ann Intern Med* 2006 Oct 03;145(7):488-496. [Medline: [17015866](https://pubmed.ncbi.nlm.nih.gov/17015866/)]
  11. Bastakoti M, Muhailan M, Nassar A, Sallam T, Desale S, Fouda R, et al. Discrepancy between emergency department admission diagnosis and hospital discharge diagnosis and its impact on length of stay, up-triage to the intensive care unit, and mortality. *Diagnosis (Berl)* 2021 Jul 05. [doi: [10.1515/dx-2021-0001](https://doi.org/10.1515/dx-2021-0001)] [Medline: [34225399](https://pubmed.ncbi.nlm.nih.gov/34225399/)]
  12. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. *Scand J Trauma Resusc Emerg Med* 2019 May 08;27(1):54 [FREE Full text] [doi: [10.1186/s13049-019-0629-z](https://doi.org/10.1186/s13049-019-0629-z)] [Medline: [31068188](https://pubmed.ncbi.nlm.nih.gov/31068188/)]
  13. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014 Sep;23(9):727-731 [FREE Full text] [doi: [10.1136/bmjqs-2013-002627](https://doi.org/10.1136/bmjqs-2013-002627)] [Medline: [24742777](https://pubmed.ncbi.nlm.nih.gov/24742777/)]
  14. Singh H, Giardina TD, Forjuoh SN, Reis MD, Kosmach S, Khan MM, et al. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf* 2012 Feb;21(2):93-100 [FREE Full text] [doi: [10.1136/bmjqs-2011-000304](https://doi.org/10.1136/bmjqs-2011-000304)] [Medline: [21997348](https://pubmed.ncbi.nlm.nih.gov/21997348/)]
  15. Goyder CR, Jones CHD, Heneghan CJ, Thompson MJ. Missed opportunities for diagnosis: Lessons learned from diagnostic errors in primary care. *Br J Gen Pract* 2015 Dec;65(641):e838-e844 [FREE Full text] [doi: [10.3399/bjgp15X687889](https://doi.org/10.3399/bjgp15X687889)] [Medline: [26622037](https://pubmed.ncbi.nlm.nih.gov/26622037/)]
  16. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer-based medical-history system. *N Engl J Med* 1966 Jan 27;274(4):194-198. [doi: [10.1056/NEJM196601272740406](https://doi.org/10.1056/NEJM196601272740406)] [Medline: [5902618](https://pubmed.ncbi.nlm.nih.gov/5902618/)]
  17. Mayne JG, Weksel W, Sholtz PN. Toward automating the medical history. *Mayo Clin Proc* 1968 Jan;43(1):1-25. [Medline: [5635452](https://pubmed.ncbi.nlm.nih.gov/5635452/)]
  18. De Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCann AP. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance of clinicians. *Br Med J* 1974 Mar 02;1(5904):376-380 [FREE Full text] [doi: [10.1136/bmj.1.5904.376](https://doi.org/10.1136/bmj.1.5904.376)] [Medline: [4594585](https://pubmed.ncbi.nlm.nih.gov/4594585/)]
  19. Zakim D, Brandberg H, El Amrani S, Hultgren A, Stathakarou N, Nifakos S, et al. Computerized history-taking improves data quality for clinical decision-making: Comparison of EHR and computer-acquired history data in patients with chest pain. *PLoS One* 2021;16(9):e0257677 [FREE Full text] [doi: [10.1371/journal.pone.0257677](https://doi.org/10.1371/journal.pone.0257677)] [Medline: [34570811](https://pubmed.ncbi.nlm.nih.gov/34570811/)]
  20. Almario CV, Chey W, Kaung A, Whitman C, Fuller G, Reid M, et al. Computer-generated vs physician-documented history of present illness (HPI): Results of a blinded comparison. *Am J Gastroenterol* 2015 Jan;110(1):170-179 [FREE Full text] [doi: [10.1038/ajg.2014.356](https://doi.org/10.1038/ajg.2014.356)] [Medline: [25461620](https://pubmed.ncbi.nlm.nih.gov/25461620/)]
  21. Kostopoulou O, Lionis C, Angelaki A, Ayis S, Durbaba S, Delaney BC. Early diagnostic suggestions improve accuracy of family physicians: A randomized controlled trial in Greece. *Fam Pract* 2015 Jun;32(3):323-328. [doi: [10.1093/fampra/cmz012](https://doi.org/10.1093/fampra/cmz012)] [Medline: [25800247](https://pubmed.ncbi.nlm.nih.gov/25800247/)]
  22. Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: A randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015 Jan;65(630):e49-e54 [FREE Full text] [doi: [10.3399/bjgp15X683161](https://doi.org/10.3399/bjgp15X683161)] [Medline: [25548316](https://pubmed.ncbi.nlm.nih.gov/25548316/)]
  23. Sibbald M, Monteiro S, Sherbino J, LoGiudice A, Friedman C, Norman G. Should electronic differential diagnosis support be used early or late in the diagnostic process? A multicentre experimental study of Isabel. *BMJ Qual Saf* 2021 Oct 05 [FREE Full text] [doi: [10.1136/bmjqs-2021-013493](https://doi.org/10.1136/bmjqs-2021-013493)] [Medline: [34611040](https://pubmed.ncbi.nlm.nih.gov/34611040/)]

24. Schwitzguebel AJ, Jeckelmann C, Gavinio R, Levallois C, Benaïm C, Spechbach H. Differential diagnosis assessment in ambulatory care with an automated medical history-taking device: Pilot randomized controlled trial. *JMIR Med Inform* 2019 Nov 04;7(4):e14044 [FREE Full text] [doi: [10.2196/14044](https://doi.org/10.2196/14044)] [Medline: [31682590](https://pubmed.ncbi.nlm.nih.gov/31682590/)]
25. Harada Y, Shimizu T. Impact of a commercial artificial intelligence-driven patient self-assessment solution on waiting times at general internal medicine outpatient departments: Retrospective study. *JMIR Med Inform* 2020 Aug 31;8(8):e21056 [FREE Full text] [doi: [10.2196/21056](https://doi.org/10.2196/21056)] [Medline: [32865504](https://pubmed.ncbi.nlm.nih.gov/32865504/)]
26. Harada Y, Katsukura S, Kawamura R, Shimizu T. Efficacy of artificial-intelligence-driven differential-diagnosis list on the diagnostic accuracy of physicians: An open-label randomized controlled study. *Int J Environ Res Public Health* 2021 Feb 21;18(4):2086 [FREE Full text] [doi: [10.3390/ijerph18042086](https://doi.org/10.3390/ijerph18042086)] [Medline: [33669930](https://pubmed.ncbi.nlm.nih.gov/33669930/)]
27. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: An online mobile survey. *J Med Internet Res* 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: [10.2196/12422](https://doi.org/10.2196/12422)] [Medline: [30907742](https://pubmed.ncbi.nlm.nih.gov/30907742/)]
28. Aaronson E, Jansson P, Wittbold K, Flavin S, Borczuk P. Unscheduled return visits to the emergency department with ICU admission: A trigger tool for diagnostic error. *Am J Emerg Med* 2020 Aug;38(8):1584-1587. [doi: [10.1016/j.ajem.2019.158430](https://doi.org/10.1016/j.ajem.2019.158430)] [Medline: [31699427](https://pubmed.ncbi.nlm.nih.gov/31699427/)]
29. Singh H, Khanna A, Spitzmueller C, Meyer AND. Recommendations for using the Revised Safer Dx Instrument to help measure and improve diagnostic safety. *Diagnosis (Berl)* 2019 Nov 26;6(4):315-323. [doi: [10.1515/dx-2019-0012](https://doi.org/10.1515/dx-2019-0012)] [Medline: [31287795](https://pubmed.ncbi.nlm.nih.gov/31287795/)]
30. Al-Mutairi A, Meyer AND, Thomas EJ, Etchegaray JM, Roy KM, Davalos MC, et al. Accuracy of the Safer Dx Instrument to identify diagnostic errors in primary care. *J Gen Intern Med* 2016 Jun;31(6):602-608 [FREE Full text] [doi: [10.1007/s11606-016-3601-x](https://doi.org/10.1007/s11606-016-3601-x)] [Medline: [26902245](https://pubmed.ncbi.nlm.nih.gov/26902245/)]
31. Davalos MC, Samuels K, Meyer AND, Thammasitboon S, Sur M, Roy K, et al. Finding diagnostic errors in children admitted to the PICU. *Pediatr Crit Care Med* 2017 Mar;18(3):265-271. [doi: [10.1097/PCC.0000000000001059](https://doi.org/10.1097/PCC.0000000000001059)] [Medline: [28125548](https://pubmed.ncbi.nlm.nih.gov/28125548/)]
32. Searns JB, Williams MC, MacBrayne CE, Wirtz AL, Leonard JE, Boguniewicz J, et al. Handshake antimicrobial stewardship as a model to recognize and prevent diagnostic errors. *Diagnosis (Berl)* 2021 Aug 26;8(3):347-352. [doi: [10.1515/dx-2020-0032](https://doi.org/10.1515/dx-2020-0032)] [Medline: [33112779](https://pubmed.ncbi.nlm.nih.gov/33112779/)]
33. Bergl PA, Taneja A, El-Kareh R, Singh H, Nanchal RS. Frequency, risk factors, causes, and consequences of diagnostic errors in critically ill medical patients: A retrospective cohort study. *Crit Care Med* 2019 Nov;47(11):e902-e910. [doi: [10.1097/CCM.0000000000003976](https://doi.org/10.1097/CCM.0000000000003976)] [Medline: [31524644](https://pubmed.ncbi.nlm.nih.gov/31524644/)]
34. Cifra CL, Ten Eyck P, Dawson JD, Reisinger HS, Singh H, Herwaldt LA. Factors associated with diagnostic error on admission to a PICU: A pilot study. *Pediatr Crit Care Med* 2020 May;21(5):e311-e315 [FREE Full text] [doi: [10.1097/PCC.0000000000002257](https://doi.org/10.1097/PCC.0000000000002257)] [Medline: [32097247](https://pubmed.ncbi.nlm.nih.gov/32097247/)]
35. Liberman AL, Bakradze E, Mchugh DC, Esenwa CC, Lipton RB. Assessing diagnostic error in cerebral venous thrombosis via detailed chart review. *Diagnosis (Berl)* 2019 Nov 26;6(4):361-367. [doi: [10.1515/dx-2019-0003](https://doi.org/10.1515/dx-2019-0003)] [Medline: [31271550](https://pubmed.ncbi.nlm.nih.gov/31271550/)]
36. Fletcher TL, Helm A, Vaghani V, Kunik ME, Stanley MA, Singh H. Identifying psychiatric diagnostic errors with the Safer Dx Instrument. *Int J Qual Health Care* 2020 Jul 20;32(6):405-411. [doi: [10.1093/intqhc/mzaa066](https://doi.org/10.1093/intqhc/mzaa066)] [Medline: [32671387](https://pubmed.ncbi.nlm.nih.gov/32671387/)]
37. Avelino-Silva TJ, Steinman MA. Diagnostic discrepancies between emergency department admissions and hospital discharges among older adults: Secondary analysis on a population-based survey. *Sao Paulo Med J* 2020;138(5):359-367 [FREE Full text] [doi: [10.1590/1516-3180.0471.R1.05032020](https://doi.org/10.1590/1516-3180.0471.R1.05032020)] [Medline: [32935740](https://pubmed.ncbi.nlm.nih.gov/32935740/)]
38. Skinner TR, Scott IA, Martin JH. Diagnostic errors in older patients: A systematic review of incidence and potential causes in seven prevalent diseases. *Int J Gen Med* 2016;9:137-146 [FREE Full text] [doi: [10.2147/IJGM.S96741](https://doi.org/10.2147/IJGM.S96741)] [Medline: [27284262](https://pubmed.ncbi.nlm.nih.gov/27284262/)]
39. Horberg MA, Nassery N, Rubenstein KB, Certa JM, Shamim EA, Rothman R, et al. Rate of sepsis hospitalizations after misdiagnosis in adult emergency department patients: A look-forward analysis with administrative claims data using Symptom-Disease Pair Analysis of Diagnostic Error (SPADE) methodology in an integrated health system. *Diagnosis (Berl)* 2021 Nov 25;8(4):479-488. [doi: [10.1515/dx-2020-0145](https://doi.org/10.1515/dx-2020-0145)] [Medline: [33894108](https://pubmed.ncbi.nlm.nih.gov/33894108/)]
40. Cheraghi-Sohi S, Holland F, Singh H, Danczak A, Esmail A, Morris RL, et al. Incidence, origins and avoidable harm of missed opportunities in diagnosis: Longitudinal patient record review in 21 English general practices. *BMJ Qual Saf* 2021 Dec;30(12):977-985 [FREE Full text] [doi: [10.1136/bmjqs-2020-012594](https://doi.org/10.1136/bmjqs-2020-012594)] [Medline: [34127547](https://pubmed.ncbi.nlm.nih.gov/34127547/)]
41. Lim G, Seow E, Koh G, Tan D, Wong H. Study on the discrepancies between the admitting diagnoses from the emergency department and the discharge diagnoses. *Hong Kong J Emerg Med* 2017 Dec 11;9(2):78-82. [doi: [10.1177/102490790200900203](https://doi.org/10.1177/102490790200900203)]
42. Adnan M, Baharuddin K, Mohammad J, Yazid B, Ahmad M. A study on the diagnostic discrepancy between admission and discharge in Hospital Universiti Sains Malaysia. *Malays J Med Health Sci* 2021 Jan;17(1):105-110 [FREE Full text]
43. Fatima S, Shamim S, Butt AS, Awan S, Riffat S, Tariq M. The discrepancy between admission and discharge diagnoses: Underlying factors and potential clinical outcomes in a low socioeconomic country. *PLoS One* 2021;16(6):e0253316. [doi: [10.1371/journal.pone.0253316](https://doi.org/10.1371/journal.pone.0253316)] [Medline: [34129648](https://pubmed.ncbi.nlm.nih.gov/34129648/)]

44. Peng A, Rohacek M, Ackermann S, Ilsemann-Karakoumis J, Ghanim L, Messmer AS, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. *Swiss Med Wkly* 2015;145:w14121 [FREE Full text] [doi: [10.4414/smw.2015.14121](https://doi.org/10.4414/smw.2015.14121)] [Medline: [25741894](https://pubmed.ncbi.nlm.nih.gov/25741894/)]
45. Eze-Nliam C, Cain K, Bond K, Forlenza K, Jankowski R, Magyar-Russell G, et al. Discrepancies between the medical record and the reports of patients with acute coronary syndrome regarding important aspects of the medical history. *BMC Health Serv Res* 2012 Mar 26;12:78 [FREE Full text] [doi: [10.1186/1472-6963-12-78](https://doi.org/10.1186/1472-6963-12-78)] [Medline: [22448755](https://pubmed.ncbi.nlm.nih.gov/22448755/)]
46. Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: A quasi-experimental study. *BMC Med Inform Decis Mak* 2006 Apr 28;6:22 [FREE Full text] [doi: [10.1186/1472-6947-6-22](https://doi.org/10.1186/1472-6947-6-22)] [Medline: [16646956](https://pubmed.ncbi.nlm.nih.gov/16646956/)]
47. Henderson EJ, Rubin GP. The utility of an online diagnostic decision support system (Isabel) in general practice: A process evaluation. *JRSM Short Rep* 2013 May;4(5):31 [FREE Full text] [doi: [10.1177/2042533313476691](https://doi.org/10.1177/2042533313476691)] [Medline: [23772310](https://pubmed.ncbi.nlm.nih.gov/23772310/)]
48. Cheraghi-Sohi S, Alam R, Hann M, Esmail A, Campbell S, Riches N. Assessing the utility of a differential diagnostic generator in UK general practice: A feasibility study. *Diagnosis (Berl)* 2021 Feb 23;8(1):91-99. [doi: [10.1515/dx-2019-0033](https://doi.org/10.1515/dx-2019-0033)] [Medline: [32083441](https://pubmed.ncbi.nlm.nih.gov/32083441/)]
49. Delaney BC, Kostopoulou O. Decision support for diagnosis should become routine in 21st century primary care. *Br J Gen Pract* 2017 Nov;67(664):494-495 [FREE Full text] [doi: [10.3399/bjgp17X693185](https://doi.org/10.3399/bjgp17X693185)] [Medline: [29074677](https://pubmed.ncbi.nlm.nih.gov/29074677/)]
50. Kostopoulou O, Sirota M, Round T, Samaranyaka S, Delaney BC. The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. *Med Decis Making* 2017 Jan;37(1):9-16 [FREE Full text] [doi: [10.1177/0272989X16644563](https://doi.org/10.1177/0272989X16644563)] [Medline: [27112933](https://pubmed.ncbi.nlm.nih.gov/27112933/)]
51. Graber MA, VanScoy D. How well does decision support software perform in the emergency department? *Emerg Med J* 2003 Sep;20(5):426-428 [FREE Full text] [doi: [10.1136/emj.20.5.426](https://doi.org/10.1136/emj.20.5.426)] [Medline: [12954680](https://pubmed.ncbi.nlm.nih.gov/12954680/)]
52. Meyer FML, Filipovic MG, Balestra GM, Tisljar K, Sellmann T, Marsch S. Diagnostic errors induced by a wrong a priori diagnosis: A prospective randomized simulator-based trial. *J Clin Med* 2021 Feb 18;10(4):826 [FREE Full text] [doi: [10.3390/jcm10040826](https://doi.org/10.3390/jcm10040826)] [Medline: [33670489](https://pubmed.ncbi.nlm.nih.gov/33670489/)]
53. Harada Y, Katsukura S, Kawamura R, Shimizu T. Effects of a differential diagnosis list of artificial intelligence on differential diagnoses by physicians: An exploratory analysis of data from a randomized controlled study. *Int J Environ Res Public Health* 2021 May 23;18(11):5562 [FREE Full text] [doi: [10.3390/ijerph18115562](https://doi.org/10.3390/ijerph18115562)] [Medline: [34070958](https://pubmed.ncbi.nlm.nih.gov/34070958/)]
54. Brandberg H, Sundberg CJ, Spaak J, Koch S, Zakim D, Kahan T. Use of self-reported computerized medical history taking for acute chest pain in the emergency department - The Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS): Prospective cohort study. *J Med Internet Res* 2021 Apr 27;23(4):e25493 [FREE Full text] [doi: [10.2196/25493](https://doi.org/10.2196/25493)] [Medline: [33904821](https://pubmed.ncbi.nlm.nih.gov/33904821/)]

## Abbreviations

**AI:** artificial intelligence

**KAKENHI:** Grants-in-Aid for Scientific Research

**OR:** odds ratio

*Edited by C Lovis; submitted 26.11.21; peer-reviewed by D Zakim, J Carter; comments to author 03.12.21; revised version received 11.12.21; accepted 02.01.22; published 27.01.22*

*Please cite as:*

Kawamura R, Harada Y, Sugimoto S, Nagase Y, Katsukura S, Shimizu T  
*Incidence of Diagnostic Errors Among Unexpectedly Hospitalized Patients Using an Automated Medical History-Taking System With a Differential Diagnosis Generator: Retrospective Observational Study*  
*JMIR Med Inform* 2022;10(1):e35225

URL: <https://medinform.jmir.org/2022/1/e35225>

doi: [10.2196/35225](https://doi.org/10.2196/35225)

PMID:

©Ren Kawamura, Yukinori Harada, Shu Sugimoto, Yuichiro Nagase, Shinichi Katsukura, Taro Shimizu. Originally published in *JMIR Medical Informatics* (<https://medinform.jmir.org>), 27.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.