BMC Medical Education

RESEARCH

Open Access

Appropriate semantic qualifiers increase diagnostic accuracy when using a clinical decision support system: a randomized controlled trial

Yasutaka Yanagita[1*], Kiyoshi Shikino[1,2], Daiki Yokokawa[1], Kosuke Ishizuka[1,3], Tomoko Tsukamoto[1], Yu Li[1], Takanori Uehara[1] and Masatomi Ikusaka[1]

Abstract

Background The role of appropriate semantic qualifiers (SQs) in the effective use of a clinical decision support system (CDSS) is not yet fully understood. Previous studies have not investigated the input. This study aimed to investigate whether the appropriateness of SQs modified the impact of CDSS on diagnostic accuracy among medical students.

Methods For this randomized controlled trial, a total of forty-two fifth-year medical students in a clinical clerkship at Chiba University Hospital were enrolled from May to December 2020. They were divided into the CDSS (CDSS use; 22 participants) and control groups (no CDSS use; 20 participants). Students were presented with ten expert-developed case vignettes asking for SQs and a diagnosis. Three appropriate SQs were established for each case vignette. The participants were awarded one point for each SQ that was consistent with the set SQs. Those with two or more points were considered to have provided appropriate SQs. The CDSS used was the Current Decision Support®. We evaluated diagnostic accuracy and the appropriateness of SQ differences between the CDSS and control groups.

Results Data from all 42 participants were analyzed. The CDSS and control groups provided 133 (60.5%; 220 answers) and 115 (57.5%; 200 answers) appropriate SQs, respectively. Among CDSS users, diagnostic accuracy was significantly higher with appropriate SQs compared to inappropriate SQs ($\chi^2(1)=4.97$, $p=0.026$). With appropriate SQs, diagnostic accuracy was significantly higher in the CDSS group compared to the control group ($\chi^2(1)=1.16\times10$, $p<0.001$). With inappropriate SQs, there was no significant difference in diagnostic accuracy between the two groups ($\chi^2(1)=8.62\times10^{-2}$, $p=0.769$).

Conclusions Medical students may make more accurate diagnoses using the CDSS if appropriate SQs are set. Improving students' ability to set appropriate SQs may improve the effectiveness of CDSS use.

Trial registration This study was registered with the University Hospital Medical Information Network Clinical Trials Registry on 24/12/2020 (Unique trial number: UMIN000042831).

Keywords Clinical decision support system, Diagnostic accuracy, Semantic qualifier

*Correspondence:
Yasutaka Yanagita
y.yanagita@gmail.com
Full list of author information is available at the end of the article

Background

The availability of medical information is increasing, and accessibility to appropriate information resources is in demand. Accordingly, the National Model Core Curriculum in Japan for undergraduate medical education (2022 revised edition) has incorporated a new ability and quality into the curriculum: being able to use information science and technology, such as artificial intelligence (AI) [1]. In this context, clinical decision support systems (CDSSs) are gaining attention as tools for mitigating diagnostic errors. A function of the CDSS is to present a list of potential diagnoses and relevant information (e.g., symptoms and medical history) when patient information is input on a specific condition [2]. The information presented by the CDSS reflects relevant guidelines and expert opinions, and information quality is generally assured [3]. Databases constructed with AI technology have also been developed. Some systems utilize these databases to display appropriate differential diagnosis lists as professionals or patients input symptoms, age, and sex, improving diagnostic accuracy [4]. CDSSs play a significant role in real-world clinical practice [5, 6] because they can help novice practitioners recall potential diagnoses and prevent diagnostic errors [7].

Although diagnostic accuracy may increase through CDSS use [8], its usefulness can only be maximized through appropriate input into the system. Thus, CDSS use effectiveness may require professionals to set up relevant semantic qualifiers (SQs), namely, keyword inputs that abstract the patient's complaints [9]. SQs are medical terms that categorize and replace specific medical history details based on higher-level concepts [9] and are often dichotomous in their logic (e.g., acute vs. chronic, local vs. radiating). While SQs are critical for structuring diagnostic reasoning and enhancing communication among health professionals, their role is equally significant in conventional diagnostic processes that lack digital support. However, few studies have addressed the comparative impact of SQs in digital and traditional diagnostic settings. Specifically, SQ selection depends on clinical abilities and supports professionals in recalling diseases, narrowing down the options from a differential disease list, and conducting related literature searches. Various researchers have evaluated CDSS accuracy and experts' diagnostic accuracy when using CDSSs [8, 11, 12]; however, despite the apparent relevance of SQs to CDSS use, few studies have examined the role and use of keywords and SQ input related to CDSS use.

Scholars have also discussed the stage of the diagnostic process at which systems such as a CDSS should be used. Meanwhile, the contribution of history-taking to diagnosis is about 80% [9], and the early-stage recalling of appropriate differential diagnoses during history-taking helps avoid diagnostic errors [13]. The diseases recalled during history-taking also affect physical examination selection and interpretation [14]. In our previous study, medical students' CDSS use led to an increased diagnostic accuracy for common diseases in an educational setting [8]. However, it remains unclear how the quality of the input data, particularly the appropriateness of SQs, influences diagnostic outcomes. CDSS relies on the input of SQs that accurately reflect the characteristics of a disease to provide reliable diagnostic suggestions. In this study, we aimed to address this gap by investigating the extent to which the appropriateness of SQs affects diagnostic accuracy when using CDSS, emphasizing that the effectiveness of CDSS depends not only on its use but also on the quality of its input. SQ appropriateness is defined as the extent to which the SQs provided by participants align with those that are both diagnostically useful and commonly established by expert physicians.

Methods

Study Design Overview

This was a randomized controlled trial study. Participants were divided into an intervention group that used the CDSS (CDSS group) and a control group that did not use the CDSS (control group). They were assigned to the groups in one clinical clerkship group unit by simple randomization using Microsoft Excel 2019 (Microsoft Corp.). Overall, 22 participants were allocated to the CDSS group and 20 to the control group. The allocation was not blinded to the participants or faculty. The participants were presented with ten case vignettes online (Microsoft Forms [Microsoft Corp.]) asking for SQs and a diagnosis. The attending physicians pre-defined the appropriate SQs and subsequently rated SQs as appropriate or inappropriate according to their agreement with the SQs previously defined as appropriate. The correct diagnosis of the case vignettes using the CDSS for SQ appropriateness and inappropriateness were compared (Fig. 1). This study was conducted in accordance with the CONSORT 2010 statement [15].

The primary outcome was the diagnostic accuracy in the CDSS group compared between appropriate and inappropriate SQs. The secondary outcome was the diagnostic accuracy of each group according to appropriate SQs.

Study Participants

Forty-two fifth-year clinical clerkship medical students in the Department of General Medicine, Chiba University Hospital, Japan, participated in this study from May to August 2020. Medical school in Japan is a six-year curriculum, and the fifth-year students participating in this study are in their first phase of patient contact. The

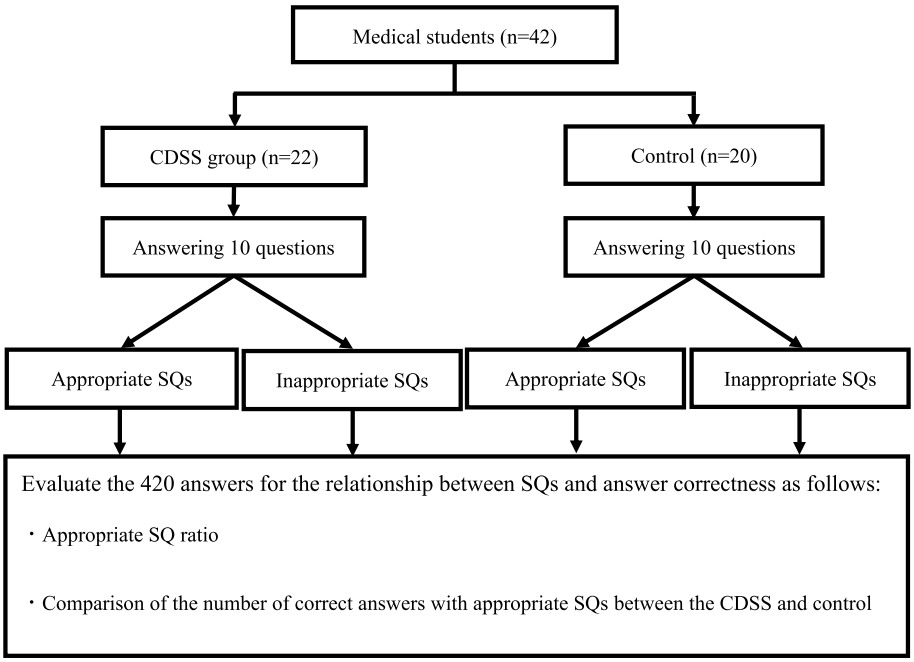

Fig. 1 Implementation flow chart. *CDSS, clinical decision support system; †SQ, semantic qualifier

participants were a representative sample of medical students in Japan. They had passed national exams conducted at medical schools across the country to evaluate student competence. These exams included shared examination to assess their knowledge and problem-solving abilities, as well as an Objective Structured Clinical Examination to assess their attitude and examination skills, thereby ensuring a certain level of competence [1]. Participants' exact exam results were considered personal information and were not used as parameters in this study. Participation was voluntary, and informed consent was obtained from all participants. Those who did not agree to participate were excluded from the study.

Experimental Materials and Procedure

The case vignettes were developed based on focus group discussions with two expert physicians (YY and KS) working in the Department of General Medicine and by referring to the diseases to be experienced as defined in the attainment goals for clinical training by the Ministry of Health, Labour and Welfare [16] and the National Medical Examination Guidelines [17]. The cases were adapted from previous research to align with the objectives of this study. The ten case vignettes were related to the field of general medicine [8]. Seven general medicine physicians with 3–7 years of medical experience were presented with the same case vignettes and were asked to answer the SQs and diagnoses without prior knowledge of the correct diagnoses. The SQs obtained from the physicians were used as references to set the appropriate SQs for each case. The difficulty level of the case vignettes was such that all seven physicians diagnosed them correctly, confirming that the difficulty level was appropriate for the participants. The appropriateness of the SQs provided by the seven physicians was independently evaluated by two experts (YY and KS). The evaluations were then compared, and in cases of disagreement, discussions were held to reach a consensus and define the three appropriate SQs essential for diagnosis in each case (Table 1). As SQ extraction is verified by focusing on diagnostic accuracy at the medical history-taking stage, the case vignettes included information on age, sex, chief complaint, and medical history and did not include content related to physical examinations or laboratory findings (Appendix 1). The CDSS used in this study was the Current Decision Support® (Precision Co., Tokyo, Japan) [18]. This system can be searched in Japanese, and when a symptom is entered, a list of differential diagnoses is displayed in categories such as high frequency and serious. Furthermore, when a disease is selected from the list, information on the disease's characteristics, diagnostic criteria, treatment methods, and drugs used are displayed. This CDSS is provided free of charge to students and faculty at the research hospital.

Microsoft Forms (Microsoft Corp.) was used to present the case vignettes, questions and answer forms. The URL for each case vignette was presented, and the participants accessed each case vignette using their own devices. In

Table 1 Case vignette lists and appropriate semantic qualifiers

Case	Diagnosis	Appropriate semantic qualifiers
1	Transient ischemic attack	Transient, weakness, dysarthria
2	Pyelonephritis	Fever, lumbar back pain, bladder irritation symptoms
3	Panic disorder	Paroxysmal attack, two sympathetic symptoms such as palpitations and lightheadedness
4	Lumbar spinal stenosis	Numbness in the lower portion of both legs, improvement with forward bending, intermittent claudication
5	Benign paroxysmal positional vertigo	Paroxysmal attack, vertigo, changes in one's head posture
6*	Acute sinusitis	Antecedent infection, facial pain, nasal discharge
7*	Acute epiglottitis	Acute, sore throat, drooling
8	Cholecystitis	Fever, postprandial symptoms, right upper quadrant pain
9	Pulmonary thromboembolism	Acute, dyspnea, unilateral leg edema
10	Acute angle-closure glaucoma	Acute, headache, visual abnormality

* Cases 6 and 7 refer to Samples 1 and 2 in the Appendix 1

the study's protocol, participants were first instructed to respond to the SQs and then proceed to provide the most probable diagnosis for each case. The answers were provided as free text, with one diagnosis and no restrictions on the number of SQs. In the intervention group utilizing the CDSS, participants followed the same initial step of answering the SQ. Subsequently, these SQs were entered into the CDSS. Based on the information obtained from the CDSS, participants then formulated and submitted their diagnoses. If the SQs provided by participants matched at least two of the three appropriate SQs for each case, the SQs were evaluated as appropriate. Two general medicine expert physicians (DY and KI) assessed SQ appropriateness. The evaluators were blinded to the groups. To evaluate the inter-rater reliability of the SQ appropriateness assessment, Cohen's Kappa was calculated to assess the agreement between the two evaluators (DY and KI). In case of a discrepancy between the two evaluators, SQ appropriateness was determined using the evaluation with the lower score. Specifically, if Evaluator A assessed that there were two appropriate SQs, but Evaluator B assessed that there was only one appropriate SQ, based on Evaluator B's assessment, the SQ setting was considered inappropriate.

Statistical Analysis

All statistical analyses were performed using SPSS Statistics for Windows 29.0 (IBM Corp.). Statistical significance was set at less than 5%. SQs to all case vignettes were classified as appropriate or inappropriate, and diagnostic correctness and incorrectness for each group were evaluated using a chi-square test. Using G*Power and considering a two-sided significance level of 5%, a power of 95%, and an effect size of 0.3, measured by Cohen's w, we estimated that 145 observations were necessary for the chi-square tests [19].

Ethics Approval

This research was performed in accordance with the Declaration of Helsinki and approved by the Ethics Review Committee of the Chiba University Graduate School of Medicine (Chiba, Japan) on May 7, 2019 (approval number: 3425). The researchers verbally obtained informed and voluntary consent from all participants. Participants were also informed that the obtained data would not be used for university grading and agreed not to share the case vignettes with other participants.

Results

All 42 eligible medical students participated in this study, and there were no dropouts. They were divided into the CDSS (CDSS use; 22 participants) and control groups (no CDSS use; 20 participants). There were 34 (81%) male participants, of which 17 (77%) were in the CDSS group and 17 (85%) in the control group. The age range of the control group was 22–32 years (standard deviation (SD): 2.9), while that of the CDSS group was 22–40 years (SD: 4.3). The mean age of the two groups was 24 years. All participants responded to all ten cases, obtaining 420 diagnoses and 1437 SQs.

Appropriate Semantic Qualifier Ratio

The mean number of SQs for each case vignette was 3.1 (SD: 0.5) in the CDSS group and 3.7 (SD: 0.3) in the control group. Regarding SQ appropriateness, the two assessors agreed on 393 of the 420 diagnoses and disagreed on 27 (κ value: 0.86). In the CDSS group, discrepancies were noted in nine diagnoses, and in the control group, 18 diagnoses. Of the 393 diagnoses on which the assessors agreed, 248 were deemed to have appropriate SQs, and 145 were deemed to have inappropriate SQs.

Table 2 The number of appropriate semantic qualifiers for the case vignettes in each group

Group	Appropriate semantic qualifiers	Inappropriate semantic qualifiers	Total
CDSS* n, (%)	133 (60.5%)	87 (39.5%)	220
Control n, (%)	115 (57.5%)	85 (42.5%)	200
Total	248	172	420

* CDSS, clinical decision support system

Table 3 The number of correct and incorrect diagnoses considering appropriate and inappropriate semantic qualifiers in the clinical decision support system group

Semantic qualifier	Correct	Incorrect	Total
Appropriate semantic qualifier n, (%)	93 (69.9%)	40 (30.1%)	133
Inappropriate semantic qualifier n, (%)	48 (55.2%)	39 (44.8%)	87
Total	141	79	220

$\chi^2(1) = 4.07$, $p = 0.026$

Table 4 Results for appropriate semantic qualifiers by diagnosis correctness and group

Group	Correct	Incorrect	Total
CDSS* n, (%)	93 (69.9%)	40 (30.1%)	133
Control n, (%)	56 (48.7%)	59 (51.3%)	115
Total	149	99	238

$\chi^2(1) = 1.16 \times 10$, $p < 0.001$ *CDSS, clinical decision support system

In the 220 diagnoses analyzed for the CDSS group, there were 133 (60.5%) and 87 (39.5%) appropriate and inappropriate SQs, respectively; in the 200 diagnoses of the control group, these numbers were 115 (57.5%) and 85 (42.5%; Table 2), respectively. The proportion of appropriate SQ settings did not differ between the groups ($\chi^2(1) = 0.38$, $p > 0.05$). The diagnostic accuracy in the CDSS group was compared between appropriate and inappropriate SQs. Table 3 presents the analysis of the primary outcome, comparing diagnostic accuracy between appropriate and inappropriate SQs.

In the CDSS group, there were 141 correct diagnoses (64.1%; out of 220 diagnoses), among which 93 (69.9%) had appropriate SQs. Among CDSS users, diagnostic accuracy was significantly higher for appropriate SQs compared to that for inappropriate SQs ($\chi^2(1) = 4.97$, $p = 0.026$; Table 3). In the control group, there were 101 correct diagnoses (50.5%; out of 200 diagnoses), among which 56 (48.7%) had appropriate SQs. Table 4 presents the secondary outcome, examining differences in diagnostic accuracy for appropriate SQs between the CDSS and control groups. With appropriate SQs, diagnostic accuracy was significantly higher in the CDSS group compared to the control group ($\chi^2(1) = 1.16 \times 10$, $p < 0.001$), as the number of correct diagnoses was significantly higher than in the control group (Table 4). Further, there was no significant difference in diagnostic accuracy between the CDSS and control groups ($\chi^2(1) = 8.62 \times 10^{-2}$, $p = 0.769$), and no significant difference in inappropriate SQs was found (Table 5).

Table 5 Results for inappropriate semantic qualifiers by diagnosis correctness and group

Group	Correct	Incorrect	Total
CDSS* n, (%)	48 (55.2%)	39 (44.8%)	87
Control n, (%)	45 (52.9%)	40 (47.1%)	85
Total	93	79	172

$\chi^2(1) = 8.62 \times 10^{-2}$, $p = 0.769$ *CDSS, clinical decision support system

Discussion

In this study, we investigated whether the keywords (i.e., SQs) input into the CDSS contribute to diagnostic accuracies among medical students. The results of this study suggest that using a CDSS in which users enter symptoms and SQs, can improve diagnostic accuracy if the SQs are entered appropriately. However, upon inputting inappropriate SQs into the CDSS, diagnostic accuracy did not increase in the CDSS group when compared to no CDSS use. This study thus reveals that SQs are important factors for CDSS use. This corroborates prior research that shows that input data is an influential process in CDSS implementation [20].

These findings emphasize that it may be crucial for medical educators to improve medical students' ability to convert patient symptoms into medical terminology that can be used in information technology tools. The analysis showed that SQ appropriateness modified the impact of CDSS on diagnostic accuracy. This highlights the role of SQs as effect modifiers in the diagnostic process, especially in CDSS. Additionally, in both groups, the proportion of appropriate SQs was higher than that of correct diagnoses. This indicates that, although participants could extract appropriate SQs from medical history information, they may not have been able to properly evaluate the recalled diseases and the list of diseases presented in a CDSS. Moreover, failed disease recall may stem from insufficient knowledge, an inability to perform proper SQs transformations, or difficulty in evaluating the list of suggested diseases from the CDSS.

When there was a difference between two evaluators, the appropriateness of the SQ was determined using the lower score. We recognize that this criterion is a stricter standard for extracting appropriate SQs. However, there is also a possibility that this procedure introduces bias,

classifying more SQs as inappropriate. Evaluators agreed on 393 of 420 diagnoses with a kappa value of 0.86, noting discrepancies in 9 and 18 diagnoses in the CDSS and control groups, respectively. The variation in medical students' SQ responses may have led to differing interpretations by evaluators.

The literature shows that the CDSS increases diagnostic accuracy [21, 22]. In another study to validate CDSS accuracy, medical student participants' diagnostic accuracy improved considerably with CDSS use. Medical students who used a CDSS achieved diagnostic accuracy comparable to residents who did not use it, particularly for high-frequency diseases, which are defined as commonly encountered in general medical practice [8]. Previous research has shown that the current challenge with CDSS use is its reliance on user ability. In many reports on CDSS usefulness, the medical professionals using a system are experts who consciously or unconsciously know which symptoms should be entered as SQs, implying that a CDSS is not intended for use by medical professionals with inadequate knowledge [23]. In this study, the participants were asked to set their own SQs without any devices. If a CDSS were evaluated with a function to assist with input or a design that can handle synonyms and colloquial expressions, the system could also be involved in setting the appropriate SQs, and different results may have been obtained. Accordingly, medical curricula should emphasize the importance of SQs when using a CDSS. Additionally, integrating CDSS training with clinical reasoning education may ensure that students develop both the technical and cognitive skills necessary for effective diagnostic processes. Combining these aspects may help medical students use CDSSs more effectively while improving their overall diagnostic abilities. In other words, SQ selection and input appropriateness are related to the user's knowledge and experience in medicine. Scholars are thus urged to further investigate user interpretation and evaluation of the information output by CDSS through SQ input. Therefore, in the current situation of clinical practice for beginner practitioners and residents and the demand for appropriate CDSS use, it may be essential to incorporate learning about SQs that represent a disease into medical curricula. Specifically, students could learn about SQs and disease concepts simultaneously. This may help improve diagnostic accuracy among residents and novice physicians, motivate them to use a CDSS effectively, and ensure efficient use of a CDSS.

When conducting medical interviews, physicians need to elicit relevant information from patient narratives. Specifically, practitioners should form disease hypotheses based on specific complaints as early as possible and proceed with the diagnostic process according to these hypotheses [24]. Considering these procedures and their common use in clinical practice, if practitioners can convert patients' ambiguous complaints into appropriate SQs—even if the complaints are input as SQs as described by patients—it may be feasible to use a CDSS to automatically generate a list of differential diseases. Furthermore, understanding the relevance of a symptom with the disease is important. Adding irrelevant or unrelated symptoms to the system could alter the list of differential diagnoses and lead physicians in the wrong direction. In addition, the quality of CDSS output may vary greatly depending on the algorithm and data source that form the basis of the system. These considerations emphasize the necessity for residents and novice clinicians to perform appropriate transformations of SQs and to possess the ability to evaluate the medical information output by the CDSS.

Limitations

This study has several limitations. First, it was conducted at one institution with fifth-year students who were part of a six-year medical school course. This study assumed that all participants had surpassed a basic level of diagnostic ability, as measured by the shared examination, which all participants passed. However, individual differences in SQ formulation skills were not explicitly analyzed, and the potential impact of exceptionally high-performing participants was not considered. In future research, it will be necessary to conduct preliminary tests to ascertain the abilities of participants and investigate how individual differences in skill level, including those of exceptionally high-performing participants, affect the results.

Second, we did not set a limit on the number of SQs that could be provided; this means that even if participants provided numerous inappropriate SQs, we still considered their SQs appropriate if they included two of three previously defined SQs. This may have influenced the accuracy of the findings. Furthermore, in this study, three appropriate SQs were set by the expert physicians following previous studies, but the appropriate SQs and the number of SQs may vary depending on the case. Depending on the number of SQs, the accuracy of evaluating the appropriate SQs and the diagnosis could be affected. There are no standardized SQ setting guidelines, because even for the same disease, SQ differs depending on the typicality of the case (e.g. typical case vs atypical case) [25]. Additionally, it can be difficult to convert symptoms into SQs because patients may describe the same symptoms in different ways.

Third, the case vignettes used in this study were paper cases and included brief medical history information written by physicians using medical terminology.

These were all common and typical vignettes, including some SQs, which may have made it relatively easy for participants to select and define the appropriate SQs. Therefore, the cases in this study do not fully replicate real-world scenarios where patients describe their complaints in their own words, which may present greater challenges in setting SQs. This warrants further research with standardized patients or real clinical settings to evaluate SQ setting under realistic conditions. Additionally, atypical cases, rare cases, and cases with complex contexts in which diagnostic errors are most often seen have not been evaluated. Thus, future researchers should verify whether practitioners can devise appropriate and acceptable SQs based on variable conditions such as patients' narratives, atypical medical history information, and rare diseases.

Fourth, this study assumed independence between each diagnosis in the statistical analysis. However, as each student evaluated multiple cases (10 cases per student), the data had a hierarchical structure. This analysis did not take into account this potential clustering effect. Therefore, future studies should consider using mixed-effects logistic regression models and other appropriate methods.

Conclusion

This study shows that fifth-year medical students were able to make the most probable diagnoses using the CDSS if they used appropriate SQs. This indicates the importance of improving practitioners'ability to set up appropriate SQs and critically evaluate CDSS outputs based on the SQs to enhance diagnostic accuracy.

Abbreviations

AI	Artificial intelligence
CDSS	Clinical decision support system
SQ	Semantic qualifier

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07294-5.

Supplementary Material 1

Acknowledgements

The authors thank Prof. Daisuke Sato (Fujita Health University Graduate School of Medicine) for advice on statistical analysis.

Authors' contributions

YY, KS, and MI designed and coordinated this study. YY, KS, DY, and KI collected and analyzed the data. YY, KS, and MI drafted the manuscript. TT, YL, KN, and UT revised the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work and for ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

Not applicable.

Data availability

The case vignettes used in this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This research was performed in accordance with the Declaration of Helsinki and approved by the Ethics Review Committee of the Chiba University Graduate School of Medicine (Chiba, Japan) on May 7, 2019 (approval number: 3425). The researchers verbally obtained informed and voluntary consent from all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of General Medicine, Chiba University Hospital, Chiba, Japan. [2]Department of Community-Oriented Medical Education, Chiba University School of Medicine, Chiba, Japan. [3]Department of General Medicine, Yokohama City University School of Medicine, Kanagawa, Japan.

Received: 2 April 2024 Accepted: 6 May 2025
Published online: 15 May 2025

References

1. Medical Education Model Core Curriculum Coordination Committee, Medical Education Model Core Curriculum Expert Research Committee. Model core curriculum for medical education. AY 2022 revision. 2023. Available from: https://www.mext. go.jp/b_menu/shingi/chousa/koutou/116/toushin/mext_01280. html. Accessed 25th Jan 2024.
2. Wohlgemut JM, Pisirir E, Kyrimi E, et al. Methods used to evaluate usability of mobile clinical decision support systems for healthcare emergencies: a systematic review and qualitative synthesis. JAMIA Open. 2023;6(3):ooad051.
3. Gholamzadeh M, Abtahi H, Safdari R. The application of knowledge-based clinical decision support systems to enhance adherence to evidence-based medicine in chronic disease. J Healthc Eng. 2023;2023:8550905. Published 2023 May 29. https://doi.org/10.1155/2023/8550905
4. Määttä J, Lindell R, Hayward N, et al. Diagnostic performance, triage safety, and usability of a clinical decision support system within a university hospital emergency department: Algorithm performance and usability study. JMIR Med Inform. 2023;11:e46760. Published 2023 Aug 31. https://doi.org/10.2196/46760
5. Shimizu T. System 2 diagnostic process for the next generation of physicians: "inside" and "outside" brain-The interplay between human and machine. Diagnostics (Basel). 2022;12(2):356. Published 2022 Jan 30. https://doi.org/10.3390/diagnostics12020356
6. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med. 2008;23(suppl 1):37–40.
7. Harada T, Miyagami T, Kunitomo K, Shimizu T. Clinical Decision Support Systems for Diagnosis in Primary Care: A Scoping Review. Int J Environ Res Public Health. 2021;18(16):8435. Published 2021 Aug 10. https://doi.org/10.3390/ijerph18168435
8. Yanagita Y, Shikino K, Ishizuka K, et al. Improving diagnostic accuracy using a clinical diagnostic support system for medical students during history-taking: a randomized clinical trial. BMC Med Educ. 2023;23:383. https://doi.org/10.1186/s12909-023-04370-6.
9. Chang RW, Bordage G, Connell KJ. Cognition, confidence, and clinical skills. Acad Med. 1998;73(10):S109–11.

10. Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med. 1992;156(2):163–5.
11. Diogo RCDS, Gengo And Silva Butcher RC, Peres HHC. Evaluation of the Accuracy of Nursing Diagnoses Determined by Users of a Clinical Decision Support System. J Nurs Scholarsh. 2021;53(4):519–526.
12. Kafke SD, Kuhlmey A, Schuster J, et al. Can clinical decision support systems be an asset in medical education? An experimental approach. BMC Med Educ. 2023;23(1):570. Published 2023 Aug 11. https://doi.org/10.1186/s12909-023-04568-8
13. Clark BW, Derakhshan A, Desai SV. Diagnostic errors and the bedside clinical examination. Med Clin North Am. 2018;102(3):453–64.
14. Shikino K, Ikusaka M, Ohira Y, et al. Influence of predicting the diagnosis from history on the accuracy of physical examination. Adv Med Educ Pract. 2015;6:143–8.
15. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
16. Ministry of Health, Labour and Welfare. Appendix: objectives of clinical training. Available at: https://www.mhlw.go.jp/topics/bukyoku/isei/rinsyo/keii/030818/030818b.html. Accessed April 7, 2022.
17. Ministry of Health, Labour and Welfare. The 115th national examination for medical practitioners: questions and answers. Available at: https://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryou/iryou/topics/tp210416-01.html. Accessed April 7, 2022.
18. Precision Co. 【CDS】Current Decision Support Available at: https://www.premedi.co.jp/cds/. Accessed December 25, 2024.
19. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. Behav Res Ther. 1996;28:1–11.
20. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data quality in the outpatient setting: impact on clinical decision support systems. AMIA Annu Symp Proc. 2005;2005:41–5.
21. Tang H, Ng JH. Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. BMJ. 2006;333(7579):1143–5.
22. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. Digit Med. 2020;3:17.
23. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet. 2018;392(10161):2263–4.
24. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ. 2002;324(7339):729–32. [published correction appears in BMJ. 2006 Nov;333(7575):944. Schwarz, Alan [corrected to Schwartz, Alan].
25. Nendaz MR, Bordage G. Promoting diagnostic problem representation. Med Educ. 2002;36(8):760–6.

Publisher's Note