Diagnoses supported by a computerised diagnostic decision support system versus conventional diagnoses in emergency patients (DDX-BRO): a multicentre, multiple-period, double-blind, cluster-randomised, crossover superiority trial

Wolf E Hautz, Thimo Marcin*, Stefanie C Hautz, Stefan K Schauber, Gert Krummrey, Martin Müller, Thomas C Sauter, Cornelia Lambrigger, David Schwappach, Mathieu Nendaz, Gregor Lindner, Simon Bosbach, Ines Griesshammer, Philipp Schönberg, Emanuel Plüss, Valerie Romann, Svenja Ravioli, Nadine Werthmüller, Fabian Kölbener, Aristomenis K Exadaktylos, Hardeep Singh†, Laura Zwaan†*

Summary

Background Diagnostic error is a frequent and clinically relevant health-care problem. Whether computerised diagnostic decision support systems (CDDSSs) improve diagnoses is controversial, and prospective randomised trials investigating their effectiveness in routine clinical practice are scarce. We hypothesised that diagnoses made with a CDDSS in the emergency department setting would be superior to unsupported diagnoses.

Methods This multicentre, multiple-period, double-blind, cluster-randomised, crossover superiority trial was done in four emergency departments in Switzerland. Eligible patients were adults (aged ≥18 years) presenting with abdominal pain, fever of unknown origin, syncope, or non-specific symptoms. Emergency departments were randomly assigned (1:1) to one of two predefined sequences of six alternating periods of intervention or control. Patients presenting during an intervention period were diagnosed with the aid of a CDDSS, whereas patients presenting during a control period were diagnosed without a CDDSS (usual care). Patients and personnel assessing outcomes were masked to group allocation; treating physicians were not. The primary binary outcome (false or true) was a composite score indicating a risk of reduced diagnostic quality, which was deemed to be present if any of the following occurred within 14 days: unscheduled medical care, a change in diagnosis, an unexpected intensive care unit admission within 24 h if initially admitted to hospital, or death. We assessed superiority of supported versus unsupported diagnoses in all consenting patients using a generalised linear mixed effects model. All participants who received any study treatment (including control) and completed the study were included in the safety analysis. This trial is registered with ClinicalTrials.gov (NCT05346523) and is closed to accrual.

Findings Between June 9, 2022, and June 23, 2023, 15 845 patients were screened and 1204 (591 [49·1%] female and 613 [50·9%] male) were included in the primary efficacy analysis. The median age of participants was 53 years (IQR 34–69). Diagnostic quality risk was observed in 100 (18%) of 559 patients with CDDSS-supported diagnoses and 119 (18%) of 645 with unsupported diagnoses (adjusted odds ratio 0·96 [95% CI 0·71–1·3]). 94 (7·8%) patients suffered a serious adverse event, none related to the study.

Interpretation Use of a CDDSS did not reduce the occurrence of diagnostic quality risk compared with the usual diagnostic process in adults presenting to emergency departments. Future research should aim to identify specific contexts in which CDDSSs are effective and how existing CDDSSs can be adapted to improve patient outcomes.

Funding Swiss National Science Foundation and University Hospital Bern.

Introduction

Diagnostic errors are a major health-care problem, affecting 5–15% of people who seek health care,[1] and can cause substantial harm.[1,2] These errors include incorrect, delayed, or missed diagnoses,[1] are associated with higher mortality rates than other types of medical error,[3] and result in the most common and expensive malpractice claims globally.[4] Computerised diagnostic decision support systems (CDDSSs) suggest possible diagnoses on the basis of patient signs and symptoms, and might improve diagnostic accuracy.[5,6] Although these systems have performed well in vignette studies, the results do not seem to translate to clinical practice.[7,8] Vignette studies have serious methodological flaws,[9] and most available clinical evaluations of CDDSSs are retrospective studies or have before–after designs.[7,8]

Two randomised controlled trials of decision support systems have assessed diagnostic quality as a secondary outcome.[10,11] A cluster-randomised trial in 10 663 ambulatory patients presenting to 280 primary care physicians

Lancet Digit Health 2025; 7: e136–44

See **Comment** page e108

*Contributed equally as first authors

†Contributed equally as co-senior authors

Department of Emergency Medicine, Inselspital University Hospital Bern, University of Bern, Bern, Switzerland (W E Hautz MD, T Marcin PhD, S C Hautz PhD, Prof G Krummrey MD, M Müller MD, T C Sauter MD, C Lambrigger RN, G Lindner MD, N Werthmüller MD, F Kölbener MD, Prof A K Exadaktylos MD); **Center for Educational Measurement and Faculty of Medicine, University of Oslo, Oslo, Norway** (S K Schauber PhD); **Bern University of Applied Sciences, Biel, Switzerland** (Prof G Krummrey); **Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland** (D Schwappach PhD); **Department of Medicine, University of Geneva, Geneva, Switzerland** (Prof M Nendaz MD); **Department of Emergency Medicine, Kepler Universitätsklinikum Linz, Johannes Kepler University, Linz, Austria** (G Linder, S Ravioli MD); **Insel Gruppe, Bern, Switzerland** (S Bosbach MD, P Schönberg MD, V Romann MD); **Department of Internal and Emergency Medicine, Bürgerspital Solothurn, Solothurn, Switzerland** (I Griesshammer MD, E Plüss MD); **Department of Emergency Medicine, King's College Hospital NHS Foundation Trust, Denmark**

Hill, London, UK (S Ravioli)**; Center for Innovations in Quality, Effectiveness and Safety (IQuESt), Michael E DeBakey VA Medical Center, Houston, TX, USA** (Prof H Singh MD)**; Department of Medicine, Baylor College of Medicine, Houston, TX, USA** (Prof H Singh)**; Institute of Medical Education Research Rotterdam (iMERR), Erasmus Medical Center, Rotterdam, Netherlands** (L Zwaan PhD)

Correspondence to: Dr Wolf E Hautz, Department of Emergency Medicine, Inselspital University Hospital Bern, University of Bern, 3010 Bern, Switzerland
wolf.hautz@insel.ch

Research in context

Evidence before this study

Whether the use of computerised diagnostic decision support systems (CDDSSs) in clinical practice improves diagnoses and patient outcomes is unclear. In a 2016 systematic review of studies investigating the effects of differential diagnosis generators, CDDSSs showed good accuracy of proposed diagnoses compared with the gold standard, but heterogeneity was high, and most studies were retrospective. In a subset of six studies evaluating the diagnostic quality of physicians' diagnoses before and after CDDSS consultation, small improvements were observed, but the clinical significance of these minor benefits is uncertain. A prospective clinical trial did not find significant improvements in the quality of care of ambulatory patients for whom a decision support information technology tool was used versus patients who received usual care.

We searched PubMed between Jan 1, 2015, and Jan 3, 2022, with the search terms (diagnosis OR diagnoses) AND ("decision support" OR "decision assistance" OR "decision aid") AND ("computer" OR "system" OR "electronic") for randomised controlled trials, published in any language, that evaluated general-purpose differential diagnosis tools. We identified the ELMO trial, a cluster-randomised trial involving 280 primary care physicians (general practitioners) in Belgium that included 10 663 ambulatory patients. The trial evaluated software that suggested laboratory tests for a wide range of conditions. Diagnostic error was evaluated as a secondary outcome, with no difference found between intervention and control. No differential diagnoses were suggested by the software directly. Although CDDSSs show good accuracy in differential diagnoses, their effect on patient-relevant outcomes is uncertain and data from high-quality clinical trials are scarce.

Added value of this study

To the best of our knowledge, DDX-BRO is the first randomised controlled trial to assess the effect of a CDDSS on diagnostic error in emergency medicine. This trial did not show superiority of diagnoses made with a CDDSS versus those made without. Despite the use of strict inclusion and exclusion criteria to recruit a population that was likely to benefit from the CDDSS, point estimates did not suggest a reduction in diagnostic quality risk.

Implications of all the available evidence

Retrospective and prospective clinical studies do not appear to show an improvement in the quality of diagnoses made in acute care with computerised diagnostic decision support. Effects of such tools are small at best and might not translate into patient-relevant outcomes for an unselected population. Careful evaluation of the actual benefits of CDDSS are warranted before their implementation in clinical practice. Further clinical research into the patient benefits in selected populations of novel types of diagnostic decision support is required.

(general practitioners) in Belgium evaluated a software program that suggested laboratory tests for a wide range of conditions.[10] Diagnostic error was assessed as a secondary outcome and no differences between the intervention and control (usual care) were found.[10] The other trial randomly assigned 936 ambulatory patients to a supported process of diagnosis and treatment with a general-purpose decision support tool (Problem-Knowledge Couplers) and 966 patients to an unsupported approach, and found that quality of care, including the diagnostic process, did not significantly differ between the groups.[11] However, an umbrella review of systematic reviews estimated that clinical decision support systems, most often targeted at single diseases, enhanced diagnoses by 9·9% (95% CI 4·1–19·3).[12]

Although CDDSSs have shown some benefit in vignette or before–after studies and for specific conditions, clinical evaluations of more general decision support systems have not. Consequently, uncertainty remains as to whether CDDSSs improve diagnoses or patient-relevant outcomes, and evidence from randomised trials is needed. Many diagnoses are first made in emergency departments, a setting particularly prone to diagnostic error due to the heterogeneity of presenting symptoms, the large number of patients, the resulting time pressure, and other complicating factors.[13] Therefore, any potential benefit of CDDSSs on diagnostic quality should be clearly visible in this setting. We thus conducted the differential diagnosis broadening trial (DDX-BRO) to determine whether CDDSSs improve diagnoses and clinical outcomes in patients presenting to the emergency department. We used a cluster-randomised design because introduction of a CDDSS requires technical set-up and training, which target a site and its health-care professionals rather than an individual patient. The aim of the study was to assess the effect of CDDSS use in the emergency department on diagnostic quality, diagnostic workflow, resource consumption, and patient outcomes.

Methods

Study design

This investigator-initiated, multicentre, multiple-period, cluster-randomised, crossover superiority trial was conducted at four emergency departments in Switzerland: a tertiary care centre (Inselspital Bern), a large urban hospital (Bürgerspital Solothurn), and two community hospitals (Spital Tiefenau and Spital Münsigen). Background, details of the trial design, and the statistical analysis plan have been previously published.[14,15] The study was conducted and is reported in accordance with the study protocol[16] and CONSORT for cluster randomised trials. All relevant local ethics committees and the Swiss national regulatory oversight body for

medicinal products approved the protocol. Ethics committees involved were those of canton Berne (lead committee, 2022-D0002) and of the northeast region of Switzerland. The protocol was revised once, 4 months after the start of recruitment, to elaborate on one inclusion criterion and to extend the study by one crossover period because of slow initial recruitment (appendix p 3).

The trial is registered with ClinicalTrials.gov (NCT05346523) and is closed to accrual.

Participants

Eligible patients were adults (aged ≥18 years) reporting with a primary symptom of fever of unknown origin, non-traumatic abdominal pain, syncope, or non-specific symptoms (eg, weakness, dizziness, or feeling unwell).[17] Patients were excluded if they were admitted in an acute life-threatening condition, pregnant, unable to follow the consent procedure, previously enrolled, presenting with a worsening pre-existing condition, or referred from another physician with a diagnosis. Detailed inclusion and exclusion criteria are provided in the appendix (p 3). The study relied primarily on patient-reported gender, which corresponded to legal sex in the case of all participants. The study did not collect data on race or ethnicity because most patients presenting to the emergency departments were presumed to be White; and so, recording and reporting of these data could endanger anonymity of the few patients who were not White. All patients provided written informed consent.

Randomisation and masking

Emergency departments were randomly assigned (1:1) to one of two predefined sequences of six alternating intervention or control periods and stratified by centre size (Inselspital Bern and Bürgerspital Solothurn *vs* Spital Tiefenau and Spital Münsigen). An investigator (SCH) drew one study site from box 1 (Inselspital Bern and Bürgerspital Solothurn) and another site from box 2 (Spital Tiefenau and Spital Münsigen) to allocate them to the predefined sequences. Treating physicians and study nurses on site could not be masked to the site allocation. Patients, all personnel assessing outcomes, and the trial statisticians were masked to group assignment.

Procedures

Patients presenting to a study site were first triaged as per local practice and assessed for eligibility by study nurses. Study nurses obtained written informed consent from eligible patients and—during intervention periods—prepopulated the CDDSS (Isabel Pro DDx Generator, Isabel Healthcare, UK) with the patient's demographics and primary reporting symptom. Isabel Pro was found to have the highest accuracy of suggested diagnoses among all investigated CDDSSs (pooled rate 0·89 [95% CI 0·83–0·94]; I^2=82%; p<0·001) in a 2016 systematic review and meta-analysis.[8] Consenting participants were seen by a resident physician in the order determined at triage who obtained a medical history and performed a clinical examination, including vital signs. Resident physicians imputed all signs and symptoms they deemed relevant into the prepopulated CDDSS, submitted those data, and received suggestions for possible differential diagnoses to consider from the CDDSS.

During control periods, patients were diagnosed as per local practice (usual care). During intervention periods, study nurses urged the diagnosing resident physician to query the CDDSS right after history and physical examination. After CDDSS consultation by resident physicians during intervention periods, or any time during control periods, further diagnostic testing was obtained at the physicians' discretion. During intervention periods, the CDDSS could be consulted as often as resident physicians deemed necessary. All input into and output from the CDDSS was automatically logged in the study database. CDDSS support was available only for study participants and during intervention periods. All cases in all periods were discussed by the diagnosing resident physician and emergency medicine consultant before emergency department discharge. Follow-up was scheduled for 14 (±4) days and conducted by central study staff masked to participant and site allocation.

See Online for appendix

Isabel Pro is marketed to health professionals—ie, the software is intended to support clinicians in broadening their differentials during diagnostic work-up. Users are provided with a list of potential differential diagnoses based on patient characteristics and key symptoms are entered as free text (see appendix p 28). Use of the software did not require extensive training.[18] Study nurses provided all resident physicians at participating sites with a short training course on the German interface right before first use. Important instructions for use were also provided on the interface of the software itself. Study nurses monitored CDDSS usage in all study participants and adherence to the protocol, and they reminded physicians at least twice to use the CDDSS for every patient included, if necessary.

Outcomes

The primary binary outcome (false or true) was a composite score indicating a risk of reduced diagnostic quality. The score was centrally assessed by study nurses independent of the study sites and masked to patient allocation. The score was set to true in all patients who died or obtained unscheduled medical care within 14 days, who were admitted to an intensive care unit from a ward within 24 h of presentation to the emergency department, or in whom a discrepancy was observed between the diagnosis made in the emergency department and the diagnosis at day 14. Discrepancies between diagnoses were assessed independently by two trained, masked physicians and in duplicate according to a predefined scheme (appendix p 4).[19] To validate the primary outcome, two different emergency

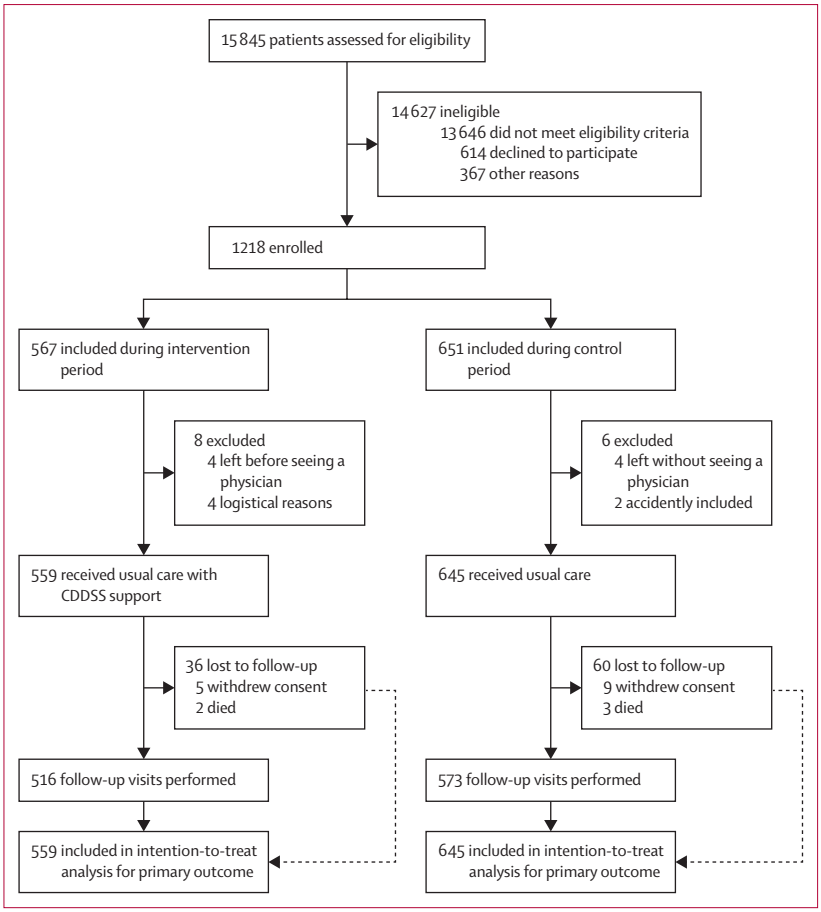

Figure: **Trial profile**
CDDSS=computerised diagnostic decision support system.

medicine consultants who were masked to patient group allocation and primary outcome performed a full record review to identify missed diagnostic opportunities using the Revised Safer Dx Instrument[20] in a random subset of 50 patients with and 50 patients without a positive diagnostic quality risk score (appendix p 5). Both measures were in fair agreement (Cohen's κ=0·5). The likelihood of identifying a missed diagnostic opportunity in patients with a positive diagnostic quality risk was 54%, compared with 4% in patients with a negative quality risk score—a 13·5 times increase.

Patients were considered lost to follow-up if they could not be reached during follow-up after three attempts on three different days and there was no additional health status information available from their primary care physician or in their hospital's electronic health record system. The primary endpoint indices for those patients were deemed to be false unless there was a change in diagnosis in a hospitalised patient or a documented revisit before the scheduled follow-up in the electronic health record system.

Secondary outcomes included all variables of the primary outcome separately, unscheduled revisits to the emergency department or primary care after 72 h and 7 days, length of stay in the emergency department, length of hospital stay if hospitalised, diagnostic tests done in the emergency department, diagnostic tests done after the emergency department visit, discharge destination, number of differential diagnoses provided by the physicians, resource consumption during the emergency department visit, and total hospital costs (incurred during a direct subsequent hospital stay or as care consumption after emergency department discharge) in Swiss francs derived from the hospitals' administrative accounting data. Prespecified safety outcomes were death, unscheduled medical care, and hospitalisation after emergency department discharge or admission to an intensive care unit other than directly from the emergency department, and were collected by central study staff. Site-investigator-adjudicated evidence of other serious adverse events and any data on CDDSS device deficiency were also collected.

The secondary outcomes of diagnostic tests done after the emergency department visit, patient-reported outcomes, and patterns of differential diagnoses provided by physicians defined in the study protocol are not reported here because they are intended to further the development of operational definitions of diagnostic quality and will be reported separately. The outcomes assessing the maximum potential benefit of CDDSS (ie, the number of cases in which the generated list of differentials included the diagnosis after 14 days) and CDDSS usage (ie, timing and number of queries) are intended for an in-depth exploration of CDDSS performance and are available only for the intervention group. These outcomes will be reported separately.

Statistical analysis

The sample size was based on the assumption of the occurrence of a diagnostic quality risk in 12% of patients in the control condition, a superiority effect of 5%, and a coefficient of variation for cluster size of 0·5. Further assumptions were a cross-sectional sampling and exchangeable correlation structure, an intra-cluster correlation between 0·01 and 0·05, and a 10% loss to follow-up. 1184 participants were required to achieve 80% power to detect CDDSS superiority at a two-sided significance level of 0·05. A sample size calculation was initially performed for four periods; however, the trial was extended to six periods during the study because of slow initial recruitment. The estimated frequency of diagnostic quality risk of 12% in the control condition was based on two previous cohort studies in emergency departments in Switzerland,[17,19] and the superiority effect was estimated from a meta-analysis of six before–after studies.[8]

The statistical analysis followed a previously published statistical analysis plan.[15] No interim analyses were conducted. The primary and all secondary outcomes were compared between groups as per intention to treat

with generalised linear mixed effects models with a binomial distribution family and exchangeable correlation structure, accounting for a random intercept for each site, resident physician, and consultant. Period (intervention or control) and patient's primary reporting symptom, age, sex, and Charlson Comorbidity Index score were factored into each analysis as covariates. Resource consumption for log-transformed Swiss francs was performed with linear models including period, site, primary reporting symptom, age, sex, Charlson Comorbidity Index, triage level, admission type, and discharge type as fixed covariates. Multiple prespecified sensitivity analyses for the primary outcome were performed, including a primary efficacy analysis, in the per-protocol sample and models without random intercepts or covariates.

We performed model checks using basic R (version 4.3.1) and the diagnostics given by R's performance package (version 0.10.4).[21] We graphically inspected normality of residuals and random effects as well as the effect of influential outliers. Furthermore, prespecified efficacy analyses were exploratively repeated for primary and secondary outcomes stratified by patient sex. Missing data were handled with full information maximum likelihood estimation because missing rates of 5% and lower were assumed to have no effect on estimates.[22] A missing rate of above 5% was not observed for any variable.

All participants who received any study treatment (including control) and completed the study were included in the safety analysis. Patients who did not complete follow-up were excluded from the safety analysis, except when reason for non-completion was death. A masked trial statistician conducted all analyses using R, version 4.3.1. A second statistician reproduced the main, per-protocol, and complete-case analysis of the primary outcome using R, version 4.2.2.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Between June 9, 2022, and June 23, 2023, 1218 patients were enrolled. 1204 patients were included in the intention-to-treat analysis: 559 in the intervention group and 645 in the control group (figure). 613 (51%) of 1204 patients were female and 591 (49%) were male (table 1). The median age of patients was 53 years (IQR 34–69). For 40 patients enrolled during an intervention period, the diagnosing physician did not use the CDDSS during the diagnostic process, and for 20 patients enrolled during a control period, the diagnosing physician self-reported usage of a different CDDSS outside of the study protocol. These patients were included in the intention-to-treat analysis but were excluded from the per-protocol analysis (appendix p 21).

	Control group (n=645)	Intervention group (n=559)
Age, years	51 (32–68)	54 (35–70)
Sex		
Male	318 (49%)	273 (49%)
Female	327 (51%)	286 (51%)
Site		
A	325 (50%)	225 (40%)
B	58 (9%)	30 (5%)
C	194 (30%)	210 (38%)
D	68 (11%)	94 (17%)
Period		
1	121 (19%)	98 (18%)
2	106 (16%)	52 (9%)
3	130 (20%)	121 (22%)
4	102 (16%)	79 (14%)
5	132 (20%)	85 (15%)
6	54 (8%)	124 (22%)
Primary symptom		
Fever of unknown origin	87 (13%)	86 (15%)
Abdominal pain	414 (64%)	340 (61%)
Syncope	70 (11%)	86 (15%)
Non-specific symptom	74 (11%)	47 (8%)
ESI triage level		
2	110 (17%)	98 (18%)
3	494 (77%)	438 (78%)
4 or 5	40 (6%)	23 (4%)
Unknown	1 (<1%)	0
Admission		
Self-admittance	577 (89%)	484 (87%)
Ambulance	64 (10%)	72 (13%)
Air rescue	0	1 (<1%)
Other	4 (1%)	2 (<1%)
Charlson Comorbidity Index		
0	279 (43%)	223 (40%)
1 or 2	173 (27%)	155 (28%)
3 or 4	120 (19%)	110 (20%)
≥5	73 (11%)	71 (13%)

Data are n (%) or median (IQR). Baseline characteristics were stratified by group allocation. ESI=Emergency Severity Index.

Table 1: **Patient baseline characteristics**

Median time from inclusion to follow-up was 13 days (IQR 12–14). The primary outcome of diagnostic quality risk was deemed to be present in 100 (18%) of 559 patients in the intervention group and 119 (18%) of 645 patients in the control group (adjusted odds ratio [OR] 0·96 [95% CI 0·71–1·30]; table 2). 64 (12%) of 559 patients in the intervention group had unscheduled medical care versus 68 (11%) of 645 patients in the control group (adjusted OR 1·21 [95% CI 0·84–1·74]). Diagnosis at emergency department discharge and diagnosis at follow-up differed in 41 (7%) of 556 patients in the intervention group

	Control group (n=645)	Intervention group (n=559)	Adjusted effect (95% CI)	p value
Primary outcome				
Diagnostic quality risk	119/645 (18%)	100/559 (18%)	OR 0·96 (0·71 to 1·30)*	0·81
Secondary outcomes				
Death within 14 days	3/645 (<1%)	2/559 (<1%)	NA	NA
Unexpected ICU admission or upscale in care within 24 h	1/645 (<1%)	0	NA	NA
Diagnostic discrepancy	60/640 (9%)	41/556 (7%)	OR 0·73 (0·47 to 1·13)*	0·16
Unscheduled medical care within 14 days†	68/645 (11%)	67/559 (12%)	OR 1·21 (0·84 to 1·74)*	0·31
Unscheduled medical care within 7 days	42/645 (7%)	55/559 (10%)	OR 1·60 (1·05 to 2·45)*	0·03
Unscheduled medical care within 3 days	29/645 (4%)	30/559 (5%)	OR 1·23 (0·72 to 2·09)*	0·45
ED costs, ₣	908 (636 to 1256)	884 (644 to 1288)	β 0·00 (–0·07 to 0·07)‡	0·99
Total hospital costs, ₣	1139 (744 to 4937)	1206 (747 to 5390)	β –0·04 (–0·11 to 0·03)†	0·10
Length of ED stay, h	4·70 (3·45 to 6·38)	4·52 (3·48 to 5·90)	IRR 0·96 (0·91 to 1·01)§	0·10
Length of hospital stay, days (hospitalised patients only)	4·0 (2·0 to 6·0)	3·0 (2·0 to 6·0)	IRR 0·93 (0·85 to 1·02)§	0·11
Blood tests performed	621/645 (96%)	533/559 (95%)	OR 0·69 (0·38 to 1·23)¶	0·21
Urine test performed	427/645 (66%)	347/559 (62%)	OR 0·87 (0·67 to 1·14)¶	0·32
Sputum test performed	124/645 (19%)	140/559 (25%)	OR 1·31 (0·93 to 1·84)¶	0·12
MRI performed	21/645 (3%)	20/559 (4%)	OR 1·05 (0·92 to 1·41)¶	0·88
CT performed	242/645 (38%)	221/559 (40%)	OR 0·97 (0·75 to 1·26)¶	0·83
Sonography performed	311/645 (48%)	257/559 (46%)	OR 0·97 (0·75 to 1·23)¶	0·84
X-ray performed	82/645 (13%)	66/559 (12%)	OR 0·84 (0·56 to 1·24)¶	0·37
Discharged home from ED	415/645 (64%)	347/559 (62%)	NA‖	NA
Differential diagnosis in addition to primary ED discharge diagnosis				
None	501/644 (78%)	444/559 (79%)	NA‖	NA
1	66/644 (10%)	63/559 (11%)	NA‖	NA
2	61/644 (10%)	42/559 (8%)	NA‖	NA
≥3	16/644 (2%)	10/559 (2%)	NA‖	NA

Data are n/N (%) or median (IQR), unless otherwise specified. β denotes unstandardised regression coefficients. ED=emergency department. ICU=intensive care unit. IRR=incidence rate ratio. NA=not applicable. OR=odds ratio. *Generalised mixed model adjusted for period (intervention or control) and patient's age, sex, primary reporting symptom, and Charlson Comorbidity Index as fixed covariates, and resident physician and site as random intercept (appendix pp 10–11). †Assessed after 14 (±4) days (see Methods section). ‡Linear model for log-transformed cost data adjusted for period, site, patient's age, sex, primary reporting symptom, Charlson Comorbidity Index, triage level, admission type, and discharge type as covariates (appendix p 12). §Generalised mixed model with a Poisson distribution for length of stay adjusted for site, period, age, sex, primary reporting symptom, Charlson Comorbidity Index, admission type, and triage category as covariates (appendix pp 13–14). ¶Primary efficacy analysis without physicians as random intercept due to model conversion failure (appendix p 15). ‖Descriptive comparison only.

Table 2: **Primary and secondary efficacy outcomes**

compared with 60 (9%) of 640 patients in the control group (adjusted OR 0·73 [95% CI 0·47–1·13]). Death within 14 days (n=5) and unscheduled intensive care unit admission within 24 h in hospitalised patients (n=1) were rare and did not significantly differ between groups (table 2). Resource consumption in Swiss francs did not differ between groups (table 2). Full model outputs and sensitivity analyses (including per protocol) are shown in the appendix (p 9).

	Control group (n=576)	Intervention group (n=518)
Patients with serious adverse events during follow-up	46 (8%)	48 (9%)
Type of serious adverse event		
Death	3 (6·5%)	1 (2%)
Hospitalisation or prolongation of existing hospitalisation	13 (28%)	20 (42%)
Life-threatening illness or injury	7 (15%)	8 (17%)
Medical or surgical intervention to prevent life-threatening illness, injury, or permanent impairment	23 (50%)	19 (40%)

Data are n (%) of safety population (n=1094). Hospitalisation refers to a stay in hospital.

Table 3: **Safety outcomes**

Serious adverse events occurred in 94 patients: 46 in the control group and 48 in the intervention group (table 3). None of the adverse events were adjudicated as possibly related to the intervention or other study procedures, and no CDDSS device deficiencies were observed.

Preplanned explorative subgroup analyses revealed a significant difference in the effect of CDDSS usage on diagnostic discrepancy between patient sexes, with a significantly lower discrepancy of diagnoses in female patients in the intervention group (adjusted OR 0·52 [95% CI 0·29–0·92]) and no difference in male patients (1·19 [0·56–2·54]; p=0·045 for group-by-sex interaction; table 4).

Discussion

This study compared diagnoses supported by a CDDSS versus conventional diagnoses in patients treated in the emergency department. Although different strategies have been used in studies to measure diagnostic quality risks, our study supports previous reports of risks related to diagnostic errors in this population of patients[1,13,19] and highlights the need to study the effect of proposed interventions to improve diagnostic accuracy.

The DDX-BRO trial was designed in several ways to demonstrate a potential benefit of CDDSS usage. First, we used strict inclusion and exclusion criteria with the aim of studying a population at high risk for diagnostic error and, hence, most likely to benefit from CDDSS usage. Although diagnostic pathways are well defined for approaching key symptoms such as chest pain or dyspnoea, the pathways for evaluating patients with less specific presenting symptoms (eg, fever of unknown origin) are less well established and many alternative diagnoses exist.[17] Because these symptoms are more

challenging to diagnose, we expected a more observable effect of CDDSS use on diagnostic accuracy in this diagnostically challenging population. Although no previous study has used the same primary outcome, one previous observational study reported a 12·8% rate of discrepancy between diagnosis at emergency department admission and diagnosis at hospital discharge in patients presenting with a wide variety of symptoms.[19] This rate of diagnostic discrepancy is 13% (58 of 442 hospitalised patients) in our trial, suggesting that our cohort included patients with a similar diagnostically challenging profile.

Second, a large body of research into the cognitive psychology of diagnostic decision making suggests that physicians generate diagnostic hypotheses early during patient encounters, and subsequent data collection, data interpretation, and final diagnoses are strongly affected by these early hypotheses.[23–26] Thus, interventions aimed at broadening the initial list of differential diagnoses should be used early in the process. This assertion is supported by a vignette study of 190 students, resident physicians, and practising physicians in emergency and internal medicine who were each asked to diagnose 16 vignette cases.[27] The study found significantly better lists of differential diagnoses when a CDDSS (Isabel Pro) was used early than when used late in the diagnostic process.[27] Thus, effects on diagnostic accuracy should be strongest when CDDSSs are used early in the diagnostic process, as in DDX-BRO. However, whether consultation of a CDDSS immediately after taking a patient history and conducting a physical examination[27] is sufficiently early or whether even earlier consultation would be optimal remains an open question.

Third, previous research has repeatedly shown low CDDSS uptake by physicians.[28] We circumvented this problem by implementing an obligation to use the CDDSS, but this might have led to pro-forma usage instead. However, data from physician questionnaires and CDDSS usage monitoring suggest that CDDSSs were almost always used as intended, and the per-protocol analysis, which accounts for non-usage, supports the main analysis. Despite these measures, we did not find a benefit of CDDSS usage on any of the assessed outcomes.

Our finding of no benefit of a CDDSS versus usual care is in line with previous prospective clinical studies that evaluated the effect of general electronic decision support on diagnoses as secondary outcomes,[10,11] but contradicts the small-to-medium benefits found in several retrospective and vignette studies.[8,12] In these retrospective and vignette studies, reasoning errors were the only possible errors that diagnosticians could make, and the findings suggest that these were partly correctable by using a CDDSS. However, in the clinic, many other causes of diagnostic error exist[13] that are unlikely to be remediable by a CDDSS. These errors include failure to detect and collect all relevant information. All relevant information is available together in retrospective designs or at once in vignettes, whereas in clinical reality, information is collected sequentially. Furthermore, in retrospective and vignette studies, one case is diagnosed at a time without disturbances, whereas in clinical practice, resident physicians often attend to several patients simultaneously and are interrupted frequently. In a hypothesis-generating preplanned analysis stratified by sex, we found a large and unexpected benefit of the CDDSS only in women. This finding might result from the more frequent presentation of notoriously difficult-to-diagnose non-specific symptoms in women than in men;[17] however, this possibility requires further investigation.

	Control group (n=645)	Intervention group (n=559)	Adjusted effect intervention: OR (95% CI)*	p value
Diagnostic quality risk	..	..	..	0·061†
Male patients	48 (15%)	51 (19%)	1·33 (0·85–2·07)	0·21
Female patients	71 (22%)	49 (17%)	0·71 (0·47–1·09)	0·12
Diagnostic discrepancy	..	..	..	0·045†
Male patients	18 (6%)	20 (7%)	1·19 (0·56–2·54)	0·65
Female patients	42 (13%)	21 (7%)	0·52 (0·29–0·92)	0·026
Unscheduled medical care	..	..	..	0·97†
Male patients	34 (11%)	33 (12%)	1·25 (0·73–2·14)	0·43
Female patients	34 (10%)	34 (12%)	1·17 (0·70–1·96)	0·55

OR=odds ratio. *OR from generalised mixed models adjusted for age, primary reporting symptom, Charlson Comorbidity Index score, and period as fixed covariates, and site, resident doctor, and consultant as random intercept. †Patient group-by-sex interaction.

Table 4: **Primary and secondary clinical efficacy outcomes stratified by patient sex**

Our study has some limitations. First, our primary outcome was a diagnostic quality risk score. The risk score indicated the presence of an adverse outcome related to a missed diagnostic opportunity at the emergency department, but did not equate to diagnostic error or diagnostic harm, which are less frequent.[13] Consequently, the intervention might have affected error or harm, which the study was not designed to detect. The score was created on the basis of a review of outcomes used in previous studies as well as expert consultation during protocol development, and was successfully validated in a randomly selected subset of patients using an established tool for evaluation of the diagnostic process.[20] Record review and the diagnostic quality risk score were only in fair agreement. However, their agreement was high compared with that observed in other studies, as the concordance between both different raters and different approaches to assess diagnostic error is notoriously low in the literature.[29] Furthermore, it seems unlikely that any inconsistency between record review and diagnostic quality risk would favour the intervention group or the control group.

Second, we did not recruit an unselected population of patients presenting to the emergency department but included patients at increased risk of diagnostic

error to maximise the chances of observing a benefit of CDDSS usage. This selection might limit the generalisability of our findings to other presenting symptoms. Published rates of diagnostic error in emergency medicine range from 6%[13] to 15% or higher.[2] This trial identified a diagnostic quality risk in 18% of patients, suggesting that the trial population was slightly more challenging to diagnose than an unselected population. Third, we evaluated the diagnostic process or patient pathways in a randomly selected subset of patients, and we did not evaluate patient experiences, only the outcome-related risks. Physicians and patients might have benefited from the use of a CDDSS in terms of test ordering or making appropriate referrals. Fourth, all patients were discussed between diagnosing resident physicians and emergency department consultants, as is common practice in all participating trial sites. Because the accuracy of diagnoses is higher when made in teams of two than when made individually,[30] the potential benefit of the CDDSS might have been reduced compared with a setting in which only a single clinician is involved in decision making. Last, although we used the most extensively and best evaluated available system, we evaluated only a single CDDSS. Therefore, our results might not generalise to all CDDSSs. In particular, novel approaches to diagnostic decision support,[31] such as providing explanations together with suggestions or establishing a structured conversation to promote reflection, might result in different effects.

Evidence from retrospective studies[8,12] and prospective clinical studies,[8,10,11] including this study, does not support the routine use of computerised diagnostic decision support to improve the quality of diagnoses made in acute care. Effects of CDDSSs are small at best and barely translate into patient-relevant outcomes in unselected patient populations in emergency medicine. Further research should evaluate the effectiveness and efficiency of CDDSSs in settings with less diagnostic collaboration, such as local health-care centres, and investigate the optimal timing of CDDSS consultation. Novel outcome measures for diagnostic quality assessment are also urgently needed.

Contributors

WEH led the overall study. WEH, TM, and SKS wrote the first draft. TM and SKS conducted the statistical analysis. WEH, TM, SCH, GK, MM, DS, MN, HS, and LZ designed the trial. TM, SCH, GK, MM, TCS, CL, GL, SB, IG, PS, EP, VR, SR, NW, FK, and AKE collected the data. All authors critically revised the manuscript. TM and SKS accessed and verified the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

WEH reports financial support from the Swiss National Science Foundation and intramural funding from the Inselspital, University Hospital of Bern; research grants from the Swiss National Science Foundation, the EU, Dräger Medical Switzerland, and Roche Diagnostics Germany; consulting fees from AO Foundation Switzerland, MDI Medical Australia, and the Swiss Institute for Postgraduate Education SIWF; and support for attending meetings or travel from Mundipharma Switzerland. SCH reports research grants from BELearn and a Centre for Health Education Scholarship at the University of Vancouver. AKE, DS, MN, and HS report financial support for the present study from the Swiss National Science Foundation. MM reports research grants from the Swiss Heart Foundation, the International Emergency Care Foundation, and Burgergemeinde. All other authors declare no competing interests.

Data sharing

De-identified individual participant data that underlie the results will be made available on request to researchers eligible to work with such data according to Swiss legislation. Eligibility will be determined by Kantonale Ethikkomission Bern. Requests to access the data can be made by sending an email together with a research plan to the corresponding author (wolf.hautz@insel.ch). To gain access, requestors will need to sign a data access agreement. Data will be available for 10 years after publication. Additional related documents, such as a data dictionary, scripts from statistical analyses, and the informed consent form, will be made available alongside the individual participant data. The study protocol[16] and the statistical analysis plan[15] have both been published.

Acknowledgments

WEH, MN, AKE, DS, and HS received a research grant from the Swiss National Science Foundation (project number 187284) and WEH received intramural funding from Bern University Hospital for this study. Trial monitoring was provided by Nicole Söll, Bern.

References

1 The National Academies of Sciences, Engineering, and Medicine. Improving diagnosis in health care. The National Academies Press, 2015.
2 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008; **121** (suppl): S2–23.
3 Zwaan L, de Bruijne M, Wagner C, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med* 2010; **170:** 1015–21.
4 Wallace E, Lowry J, Smith SM, Fahey T. The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open* 2013; **3:** e002929.
5 Graber MA, VanScoy D. How well does decision support software perform in the emergency department? *Emerg Med J* 2003; **20:** 426–28.
6 Singh H, Graber ML. Improving diagnosis in health care—the next imperative for patient safety. *N Engl J Med* 2015; **373:** 2493–95.
7 Patterson BW, Pulia MS, Ravi S, et al. Scope and influence of electronic health record-integrated clinical decision support in the emergency department: a systematic review. *Ann Emerg Med* 2019; **74:** 285–96.
8 Riches N, Panagioti M, Alam R, et al. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PLoS One* 2016; **11:** e0148991.
9 Kopka M, Feufel MA, Berner ES, Schmieding ML. How suitable are clinical vignettes for the evaluation of symptom checker apps? A test theoretical perspective. *Digit Health* 2023; **9:** 20552076231194929.
10 Delvaux N, Piessens V, Burghgraeve T, et al. Clinical decision support improves the appropriateness of laboratory test ordering in primary care without increasing diagnostic error: the ELMO cluster randomized trial. *Implement Sci* 2020; **15:** 100.
11 Apkon M, Mattera JA, Lin Z, et al. A randomized outpatient trial of a decision-support information technology tool. *Arch Intern Med* 2005; **165:** 2388–94.
12 Borges do Nascimento IJ, Abdulazeem HM, Vasanthan LT, et al. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. *Lancet Digit Health* 2023; **5:** e534–44.
13 Newman-Toker DE, Peterson SM, Badihian S, et al. Diagnostic errors in the emergency department: a systematic review. Agency for Healthcare Research and Quality (US), 2022.
14 Marcin T, Hautz SC, Singh H, et al. Effects of a computerised diagnostic decision support tool on diagnostic quality in emergency departments: study protocol of the DDx-BRO multicentre cluster randomised cross-over trial. *BMJ Open* 2023; **13:** e072649.

15 Marcin T, Müller M, Schauber SK, Hautz WE. DDX-BRO statistical analysis plan v1.0. March 27, 2023. https://zenodo.org/records/7773123 (accessed Oct 22, 2024).

16 Marcin T, Hautz WE. DDX-BRO clinical investigation protocol v1.5. Sept 28, 2023. https://zenodo.org/records/8386430 (accessed Oct 22, 2024).

17 Nemec M, Koller MT, Nickel CH, et al. Patients presenting to the emergency department with non-specific complaints: the Basel Non-specific Complaints (BANC) study. *Acad Emerg Med* 2010; **17:** 284–92.

18 Briggs JS, Fitch CJ. The ISABEL user survey. *Med Inform Internet Med* 2005; **30:** 167–72.

19 Hautz WE, Kämmer JE, Hautz SC, 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; **27:** 54.

20 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; **6:** 315–23.

21 Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. Performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Softw* 2021; **6:** 3139.

22 Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999; **8:** 3–15.

23 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; **173:** 418–25.

24 Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians' diagnostic judgments. *Med Decis Making* 2012; **32:** 831–39.

25 Kostopoulou O, Devereaux-Walsh C, Delaney BC. Missing celiac disease in family medicine: the importance of hypothesis generation. *Med Decis Making* 2009; **29:** 282–90.

26 Jain B. The key role of differential diagnosis in diagnosis. *Diagnosis (Berl)* 2017; **4:** 239–40.

27 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* 2022; **31:** 426–33.

28 Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement Sci* 2017; **12:** 113.

29 Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, Norman G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf* 2017; **26:** 104–10.

30 Hautz WE, Kämmer JE, Schauber SK, Spies CD, Gaissmaier W. Diagnostic performance by medical students working individually or in teams. *JAMA* 2015; **313:** 303–04.

31 Schmidt HG, Mamede S. Improving diagnostic decision support through deliberate reflection: a proposal. *Diagnosis (Berl)* 2022; **10:** 38–42.