BMJ Open

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041396
Article Type:	Protocol
Date Submitted by the Author:	07-Jun-2020
Complete List of Authors:	Faqar-Uz-Zaman, S. Fatima; Hospital of the Goethe University Frankfurt Surgery Centre, Department for General, Visceral and Transplant Surgery Filmann, Natalie; Institute of Biostatistics and Mathematical Modeling, Goethe-University, Frankfurt/Main, Mahkovic, Dora; Ljubljana Central Medical School von Wagner, Michael; Goethe University Frankfurt Detemble, Charlotte ; Hospital of the Goethe University Frankfurt Surgery Centre Kippke, Ulf; Hospital of the Goethe University Frankfurt Surgery Centre Marschall, Ursula; BARMER Anantharajah, Luxia ; Hospital of the Goethe University Frankfurt Surgery Centre Baumartz, Philipp; Hospital of the Goethe University Frankfurt Surgery Centre Sobotta, Paula ; Hospital of the Goethe University Frankfurt Surgery Centre Bechstein, Wolf; Hospital of the Goethe University Frankfurt Surgery Centre Schnitzbauer, Andreas; Hospital of the Goethe University Frankfurt Surgery Centre
Keywords:	ACCIDENT & EMERGENCY MEDICINE, Adult gastroenterology < GASTROENTEROLOGY, Adult surgery < SURGERY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

S. Fatima Faqar-Uz-Zaman[1], Natalie Filmann[2], Dora Mahkovic[3], Michael von Wagner[4], Charlotte Detemble[1], Ulf Kippke[1], Ursula Marschall[5], Luxia Anantharajah[1], Philipp Baumartz[1], Paula Sobotta[1], Wolf O. Bechstein[1] and Andreas A. Schnitzbauer[1]

[1] Department for General, Visceral and Transplant Surgery, Frankfurt University Hospital, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

[2] Institute of Biostatistics and Mathematical Modeling, Goethe-University Frankfurt, Germany

[3] MCL Medical Center Ljubljana, Ljubljana, Slovenia

[4] Executive Department for Medical IT-Systems and Digitalization, Frankfurt University Hospital, Germany

[5] Barmer health insurance, Germany

Corresponding author:

Prof. Dr. Andreas A. Schnitzbauer, FACS, FEBS

Department for General, Visceral and Transplant Surgery

Frankfurt University Hospital, Goethe-University Frankfurt/Main

Theodor-Stern-Kai 7

60590 Frankfurt am Main

Phone: +49-69-6301-5253

Fax: +49-69-6301-84028

Email: andreas.schnitzbauer@kgu.de

Abstract

Introduction: Occurrence of inaccurate or delayed diagnoses is a significant concern in patient care, particularly in emergency medicine, where decision-making is often constrained by high throughput and inaccurate admission diagnoses. Artificial intelligence (AI)-based diagnostic decision support system (DDSS) have been developed to enhance clinical performance by suggesting differential diagnoses to a given case, based on an integrated medical knowledge base and machine learning techniques. The purpose of the study is to evaluate the diagnostic accuracy of Ada®, an app-based diagnostic tool, and the impact on patient outcome.

Methods and analysis: The eRadaR trial is a prospective, double-blinded study with patients presenting to the emergency room (ER) with abdominal pain. At initial contact in the ER, a structured interview will be performed using the Ada-App® and both, patients and attending physicians, will be blinded to the proposed diagnosis lists until trial completion. Throughout the study, clinical data relating to diagnostic findings and types of therapy will be obtained and the follow-up until day 90 will comprise occurrence of complications and overall survival of patients. The primary efficacy of the trial is defined by the percentage of correct diagnoses suggested by Ada® compared to the final discharge diagnosis. Further, accuracy and timing of diagnosis will be compared to decision-making of classical doctor-patient interaction. Secondary objectives are complications, length of hospital stay and overall survival.

Ethics and dissemination: Ethical approval was received by the independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 including the patient information material and informed consent form. All protocol amendments must be reported to and adapted by the IEC. The results from this study will be submitted to peer-reviewed journals and reported at suitable national and international meetings.

Trial registration: German Clinical Trials Register: DRKS00019098 (registered on 29th May 2020).

Keywords: Artificial intelligence, diagnostic accuracy, emergency room diagnoses, abdominal pain

Article Summary

Strengths and weaknesses of this study

- This is the first prospective study to examine the diagnostic accuracy of an app-based diagnostic tool in an emergency room (ED) and the impact on clinical outcomes.
- The study will be conducted in a real life setting to investigate the performance in a high stress environment and to provide rational for routine clinical application.
- The double-blinded design will avoid bias regarding research findings.
- The primary limitation of an observational design is that only associations can be described, not causal relationship.

Introduction

Diagnostic errors, comprising inaccurate, delayed or missed diagnoses, are one of the major challenges in public healthcare [1]. In the recent 'Patient Safety Fact File', the World Health Organization (WHO) outlines ten crucial facts about patient safety [2]. Accordingly, adverse events are among the ten leading causes of death and disability, contributing to approximately 10% of patients harmed during hospitalization. Of note, 10% to 20% of adverse events have been quoted to be particularly related to diagnostic failure, causing more harm to patients than medication or treatment errors [3–5]. Further, false or delayed diagnoses are reported to be the most common reason for medical malpractice litigation [6]. Graber *et al.* estimated that diagnostic failures occurred in 5%-15% of cases, depending on the medical specialty with higher percentages assumed in primary care and emergency medicine [7]. Various reasons have been identified to contribute to false diagnoses. Graber concluded that cognitive slips, primarily resulting from faulty information processing and verification, and misguided situational confidence occur most frequently [8, 9].

This is especially evident in ER settings, which often have to deal with high throughputs, fast decision-making and incomplete clinical information in a disruptive environment. In particular, ER overcrowding has been identified as a serious threat to patient safety, resulting in poor clinical outcome and a significant increase in mortality [10].

Previous studies have revealed that more than 40% of admission diagnoses at first presentation to the ER are not concordant with the final diagnosis of the patient [11–13]. That means, that throughout the hospital stay, the patient experiences a change in diagnosis based on a variety of additional diagnostics and reevaluation of initial assumptions, finally leading to the correct diagnosis. In particular, approximately 30% of patients with abdominal pain, being one of the leading cause for visiting the ER, exhibit a discrepancy in diagnosis [14, 15]. In particular, misdiagnosis rate of acute appendicitis, the most frequent reason for acute abdominal pain, has largely remained unchanged over the time and is still associated with a high ratio of negative appendectomies [16]. Inaccurate diagnosing in ERs has been shown to be further associated with increased length of hospital stay, rate of consultations, healthcare cost and risk for mortality and morbidity, contributing to a serious concern to patient safety [11, 13, 17, 18]. Thus, a high degree of diagnostic accuracy can lead to an

improvement in quality of patient care. Correct admission diagnoses are crucial for a reliable triage and process management and critically influence the initial evaluation in that ER and subsequent clinical course of the patient [19].

Digital technologies and artificial intelligence (AI)-based methods have recently emerged as impressively powerful tools to empower physicians in clinical decision making and improve healthcare quality. More specifically, diagnostic decision support systems (DDSS) have demonstrated to facilitate assessment of clinical data input by using an extensive medical knowledge base [20, 21]. One version of DDSS is Ada®, an app-based AI-machine learning system which incorporates patients' symptoms and other findings into its knowledge base and intelligent technology to deliver effective healthcare [22, 23]. Based on an algorithmic pathway and driven by chief complaints, the app-based system generates a set of differential diagnosis for a given clinical case. Several studies have reported that DDSS have the potential to increase diagnostic performance, obtaining an accuracy rate of 70-96% [24, 25]. In particular, a retrospective study of rare diseases has demonstrated that Ada suggests accurate diagnoses earlier than clinical diagnoses in more than half of all cases [23].

However, application of the Ada app has not been investigated in a real-life setting, particularly in ERs, which has to deal with a high stress environment and heavy time constraints. This app-based method may be a valuable companion in triaging patients and support clinicians in making decisions more accurate and sooner by simultaneously reducing risk for medical errors. Therefore, in the present study, we aim to evaluate the diagnostic ability of Ada in ER settings and examine the impact on timing of diagnosing.

Methods and analysis

The eRadaR-trial is designed as a prospective, double-blinded, observational study evaluating the diagnostic accuracy of the Ada-App® in the ER of the Department of General, Visceral and Transplant Surgery of the Frankfurt University Hospital, Germany. This trial is registered as DRKS00019098 in the German Clinical Trials Register and the trial protocol is written in accordance with the current Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT 2013). The SPIRIT checklist is given in Additional file 1.

Study population and eligibility criteria

All patients presenting to the ER with abdominal pain will form the study population and be screened for trial eligibility. Inclusion criteria comprise: (a.) adults aged ≥ 18 years, (b.) patients presenting with abdominal pain to the ER and (c.) patients willing to participate and able to provide written informed consent. The criteria of exclusion are: (a.) intubated patients, (b.) instable patients or (c.) patients with severe injuries requiring immediate medical treatment, (d.) patients unwilling or incapable of providing informed consent. Eligible patients are asked for their participation in the trial and written informed consent will be obtained from themselves. All reasons for exclusion of patients will be recorded in the trial screening log and analyzed accordingly.

Description of study visits and assessment schedule

Eligible patients will be interviewed by the study team with the Ada-App® based on an algorithmic pathway of questions relating to the symptoms. Throughout the study, the patient, the study team, and the physician treating the patient will be blinded regarding the list of proposed diagnoses by the app. The patient will subsequently be diagnosed by classical doctor-patient interaction and decision-making. The clinical course of the patient will be followed until day 90 after initial contact in the ER. Detailed information about outline of the study and assessment schedule are displayed in Table1 and Figure1.

Table 1 Schedule of study visits and assessments of the eRadaR study

	Baseline	**Hospital stay***	**Discharge**	**90-days FU**
Visits	V1 (Day 0)	V2 and 3 (Day 7, 14)	V4	V5 (Day 90)
Informed consent	X			
Eligibility criteria	X			
Demographic data:	X			X
a) Charlson Comorbidity Index	X			
b) RAI-C score				
Ada diagnosis list	X			
ICD-10 diagnoses	X		X	X
Symptoms	X			
Diagnostics**		X	X	
Therapy and OPS-code		X	X	
Rate of consultations			X	
Complications (CCI)		X	X	X
Length of hospital stay			X	
Overall survival				X

* Visit 2 or 3 are left out, if the patient is discharged before
** Diagnostics include:
a) Routine blood samples (C-reactive protein, White blood cells, Hemoglobin, Platelets, Sodium, Potassium, Creatinine, Albumin, Bilirubin, INR)
b) Instrumental diagnostics (Ultrasound, Chest/Abdominal CT/MRI, ECG, Endoscopy)
RAI-C = Risk Analysis C score; ICD = International Classification of Diseases; OPS = Operations and Procedures;
CCI = Comprehensive Complication Index, FU = Follow-up, V = visit

Patient presenting to the ED (Visit 1)

After enrollment in the trial, a structured interview with the Ada-App® will be conducted and bassline data will be assessed including demographic data according to the Carlson Comorbidity Index and the Risk Analysis Index-C score (RAI-C score), the patients' symptoms and ICD-10 diagnoses list [26–28]. Participants are then diagnosed and treated according to the standard of care by the attending physician of the ER. As this is a double-blinded study to patients and treating physicians, Ada-App®

diagnoses lists will be randomly allocated to a study-ID and then manually transferred into the electronic case report forms (eCRF). The trial personnel will be blinded until the end of the study to avoid bias regarding subsequent diagnoses and treatment of the patient, except of the interim analysis, which is mentioned is the section of statistical analysis.

Hospital stay (Visit 2, day 7)

This visit is performed on day 7, after the patient is admitted to the hospital. Data about diagnoses and therapies are assessed comprising laboratory results (i.e. C-reactive protein, white blood cells, platelets, hemoglobin, bilirubin, creatinine, sodium and potassium, albumin, INR), computer-assisted diagnostics (i.e. ultrasound, chest/abdominal CT/MRI, ECG, endoscopy), type of therapy (conservative, interventional or surgery), OPS-code of therapies and complications according to the Comprehensive Complication Index (CCI) together with the date of occurrence [29]. If the patient has not been admitted to the hospital or is discharged before day 7, visit 2 is left out.

Hospital stay (Visit 3, day 14)

Visit 3 is performed on day 14 after patient's admission and assessment schedule is equivalent to visit 2. If the patient is discharged before day 14, visit 3 is left out.

Discharge (Visit 4)

At discharge, data including the final ICD-10 diagnosis and the timing of diagnosis will be recorded to subsequently analyze the accuracy and the timing of the Ada-App® compared to the classical doctor-patient-encounter. Further data items include diagnostics (laboratory, instrumental), OPS-codes and type of therapies, complications according to CCI, length of hospital stay, overall health cost, rate of consultation.

Follow-up (Visit 5)

The follow-up will be performed as a structured telephone interview or in person on day 90 and will encompass following data items: demographic data according to the RAI-C score, complication assessment according to CCI and overall survival.

Interventions

As this is an observational, double-blinded, prospective study, no experimental or control interventions are conducted.

Endpoints

Primary endpoint

The primary endpoint of this study is to evaluate the diagnostic accuracy of the Ada-App® by comparing the decision-making of the classical doctor-patient interaction with the diagnoses proposed by the app-based algorithm.

Secondary endpoint

Secondary endpoints of this study consist of the following: timing of final discharge diagnosis and time to treatment during hospital stay, comparing accurate diagnoses with discharge diagnoses as descriptive assessments, the occurrence of complications according to the CCI, total length of stay in hospital from initial contact in the ER until discharge, patient morbidity and mortality at day 90, overall health cost analysis and consultation rate. Further endpoints are displayed in the description of assessment schedule (Table1).

Measurement methods

For data capture, following measurement methods will be used:

1. Primary outcome measurement will be performed using the Ada-App® which will deliver a set of differential diagnoses to a given clinical case [23]. Based on an algorithmic questionnaire and machine learning technologies, the Ada chatbot assesses symptoms of the patient, similar to the anamnestic techniques and clinical reasoning of physicians. Patients' data are integrated into an extensive knowledge base, which has been specifically designed by medical doctors by incorporating validated disease models and comprehensive medical literature. Then, differential diagnoses are generated and ranked in order considering two features: the probability, based on epidemiologic data, and the best match between the diagnosis and the given symptoms. Through AI-based methods and multiple feedback loops, the Ada® knowledge base grows after each interaction and diagnostic ability improves continuously.
2. The occurrence of complications as secondary outcomes will be evaluated and analyzed according to the Comprehensive Complication Index (CCI) [29]. The CCI represents the standard assessment of postoperative morbidity and comprises all complications occurring during a patient's course based on the Clavien-Dindo classification (CDC). Compared to the CDC, which ranks complications based on

the severity of the therapeutic consequence and grades them in 5 levels, the CCI uses a formula to integrate all complications, ranging them from 0 ('no complication') to 100 ('death') [30]. This advanced approach enables comparison of patients harboring more than one complication and takes more subtle differences into consideration.

3. For assessment of comorbid diseases and frailty-associated risk in a surgical population we will use the Charlson Comorbidity Index and the RAI-C score.

Risk-Benefit assessment

This is an observational, non-interventional study and does not comprise any specific risk for the patient, as data obtained with the app are not used in the ER standard of care. Therefore, there is no special need for additional safety management. A delay in the diagnosis and treatment of patients presenting to the ER is not expected, as the app-based interview will not require more than 10 minutes and will exclusively be performed in the waiting zone of the ER by the study team. Baseline assessment (during visit 1) will directly be conducted after patient has been registered at the ER and given informed consent. Besides that, instable patients requiring immediate medical care are excluded from the study beforehand.

Data management and data safety

The investigators will design and produce electronic case report forms (eCRF) for protocol-required data collection. All information will be entered into these eCRFs by authorized and trained members of the study team and systematically checked for accuracy and completeness. Staff members with responsibilities for data collection or those, having access to the database will be enrolled in a delegation log. Patients' data collected during the trial will be recorded in pseudonymized form by solely using individual identification codes.

For data assessment using the Ada-App®, a specified iPad will be provided, which will be registered at the Frankfurt University Hospital and will be exclusively used for the purpose of this trial. Clinical data will be documented pseudonymously by using a combination of a random number from 1 to 450 and the patient's year of birth. Participants are then asked to answer the questionnaire of the Ada-App® preferably by themselves or otherwise assisted by the study team. The diagnoses will be manually transferred into the eCRF of the related patient after trial completion and unblinding.

All trial data obtained will be integrated in a statistical analysis software and analyzed by the Institute of Biostatistics and Mathematical Modeling Frankfurt.

Publication policy

The results of this trial will be submitted for publication in a peer-reviewed journal in a summarized anonymized manner. The study is scientifically supported by the Barmer health insurance company. Barmer will act as a scientific advisor regarding the conduct of the study, will be involved in the process of interpreting the data and in the publication and public distribution process of the study after trial completion. However, there will be no raw data sharing or financial support by the institution.

Statistical analysis

Interim analysis

One formal unblinded interim analysis of the trial data is planned to be performed after enrollment of about 200 patients to evaluate the diagnostic accuracy of the Ada-App® with 90-days follow-up information. Statistical analysis will be performed by the responsible study biometrician using a significance level of alpha = 0.001 and a subsequent report will be written. These results will be discussed with the investigators and the study team in a staff meeting and the continuation of the trial will be considered.

Sample size calculation and study duration

The assumptions that were made, was that more than 30% of the admission diagnoses are not consistent with the final discharge diagnosis and hypothesized that the Ada-App® will increase the diagnostic accuracy from 70% to a rate of 85%. Providing a power of 90% and a two-sided significance level of alpha 5%, a target sample size of N = 405 patients has to be recruited to detect the targeted effect. With an estimated dropout rate of 10%, we plan to recruit N = 450 patients in this trial. Furthermore, we expect the width of the confidence intervals for the diagnostic accuracy to be 0.1 at maximum (0.09 with an estimated diagnostic accuracy of 0.7, 0.07 with an estimated diagnostic accuracy of 0.85).

This trial is anticipated to start in September 2020 and the duration of patient's participation is 3 months including follow-up. To achieve the required sample size of patients, trial completion is expected to be in 12 months (August 2020).

Ethical and legal aspects, consent

The eRadaR trial will be conducted in accordance with the Declaration of Helsinki and the international conference of harmonization good clinical practice (ICH-GCP) guidelines. After a patient has been identified to meet eligibility criteria, the patient will be informed about the aim, outline and individual risk of the study and the informed consent will be given. After a sufficient period the patient can then sign the informed consent and will receive a signed copy.

Patient and Public involvement

Patients were not involved in the development of the research question or study design. They will however be involved in visit 1 and will be interviewed by the study team using the Ada-App® . Further, the follow-up (visit 5) will be performed as a telephone interview or in person with the patients for data assessment.

Discussion

Diagnostic errors have been identified as a serious threat to patient safety, leading to preventable adverse events, particularly in ERs with a disruptive environment. AI-based tools and algorithms have the potential to substantially reduce diagnostic failures, achieving high rates of diagnostic accuracy, which rivals the capability of clinicians.

A previous study provides an overview about main types of existing tools, which are classified in categories related to the targeted step of diagnostic processing [25]. Over the past few decades, a number of computerized DDSS have been developed, exhibiting promising diagnostic efficacy. Bond *et al.* evaluated four current DDSS using clinical cases from the *New England Journal of Medicine*, demonstrating that Isabel and Dxplain achieve the strongest performance [31]. Compared to former programs, second generation DDSS are far more powerful, providing more accurate suggestions with increasing complexity, while concomitantly requiring less time for diagnosing [21, 24]. This is primarily essential in an era of ER crowding, where fast and accurate triaging is necessary to prioritize critically ill patients and to optimize resource allocation [8]. Stewart *et al.* recently summarized various fields of AI application becoming relevant in the emergency medicine, including imaging, decision-making and outcome prediction [32]. In terms of triaging, a machine learning based tool efficiently predicts critical patient outcome, equivalent to the classically used Emergency Severity Index [33]. In a prospective, multi-center study, the DDSS Isabel achieved high accuracy in diagnosing patients presenting to the ER, suggesting the final discharge diagnosis in 95% of cases [34]. Another clinical decision support system has been evaluated in patients presenting with acute abdominal pain aiming to identify high risk patients for acute appendicitis [35]. Based on automated methods and an integrated risk calculator, patient data was assessed from the electronic health record (EHR) and management strategies suggested according to the risk level. Incorporation into EHR represents one of the most recent advances in the development of DDSS using 'natural language processing' techniques, which matches entered clinical data with the underlying knowledge base [36]. This might facilitate assessment of larger volumes of data, save more time and might increase acceptance of DDSS in clinical workflow.

However, in most of these trials using clinical support systems, impact on patient outcome or on healthcare costs were not assessed. Although diagnoses suggested by

DDSS mostly contained the correct diagnosis and achieved high level of users' satisfaction, relevance and specificity of extensive lists were low [21, 24, 31]. Long lists may lead to distraction or to unnecessary diagnostic with increased risk for iatrogenic injuries and costs. In general, despite the given potential efficacy of DDSS, widespread acceptance for implementation of DDSS into the routine clinical practice is evolving scarcely [37]. Studies focusing on AI-based diagnostic tools are generally designed heterogeneously and are often of poor quality, making it difficult to recommend widespread evidence-based clinical application [21, 25]. While most of the current trials demonstrated high diagnostic accuracy in retrospective and simulated cases, only few studies evaluated their performance in real clinical settings, particularly in high stress environments like ERs. Thus, further validations in prospective studies are required to investigate the diagnostic efficiency and utility of DDSS and their impact on routine clinical decision-making and patient outcome.

Trial status

Ethical approval for this trial was granted by an independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 and anticipated trial start date is September 2020.

Abbreviations

AI, Artificial Intelligence; App, Application; CCI, Comprehensive Complication Index; CDC, Clavien-Dindo classification; DRKS, Deutsches Register Klinische Studien (German Clinical Trials Register); eCRF, electronic case report form; DDSS, Diagnostic Decision Support System; ER, Emergency Room; EHR, Electronic Health Record; FU, Follow-Up; ICD-10, International Classification of Diseases; ICH-GCP, International Conference of Harmonisation, Good Clinical Practice; IEC, Independent ethics committee; OPS, Operations and Procedures; RAI-C, Risk Analysis Index-C score.

Acknowledgements

We would like to thank Rene Wendel, Doreen Giannico, Dr. Johannes Masseli, Prof. Stefan Zeuzem and the whole administrative and medical team of the central emergency room of the Frankfurt University Hospital for their support.

We further are grateful for our established scientific support with the Barmer health care insurance company.

Funding

There is no funding for this trial from third parties or grants. This is an investigator-initiated trial.

Competing interest

No competing interests were detected for the specific trial.

Author Contributions

SFF has written the manuscript, was involved in writing the protocol and is the clinical lead surgeon for the trial.

NF performed the sample size calculation for the trial and was involved in drafting the protocol and the manuscript.

MvW supported the trial as Chief Medical Informatics Officer (CMIO) of the hospital, and was involved in drafting the protocol and the manuscript

CD and DM were involved in creating the idea and drafting the protocol and the manuscript and prepared the CRF logistics.

UK is the quality manager of the surgical department and supervised implementation of the project in the ER.

UM is the BARMER health insurance representative and is an external advisor to the trial. She drafted the protocol and the manuscript.

LA, PB, PS are medical students triaging the patients and putting up logistics in the ER. They all were involved in creating the idea and shaping the project.

WOB gave valuable input into the project, supports it majorly as chair of the department and creates a culture for innovative projects. He further drafted the protocol and manuscript.

AAS had the idea for the project, leads the trial group and links all parties involved. He is the responsible principal investigator for the trial.

Availability of data and materials

Data and material will be available after an adequate research proposal has been made and reviewed by the responsible persons at the University. The requests will be forwarded to our Ethics committee and the requesting party will be responsible to get

permission to do post-hoc analysis and will be also liable for any costs arising from the requests.

Consent for publication

All authors read and consented on the publication of the current version of this manuscript.

References

1. Balla J, Heneghan C, Goyder C, Thompson M. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. BMJ Open 2012.
2. World Health Organization. Patient Safety and Risk Management Service Delivery and Safety. Accessed Sept 2019. https://www.who.int/features/factfiles/patient_safety/en/.
3. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington (DC); 2000.
4. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998;169.
5. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324.
6. Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, et al. Claims, errors, and compensation payments in medical malpractice litigation. N Engl J Med. 2006;354.
7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121.
8. Hautz SC, Schuler L, Kämmer JE, Schauber SK, Ricklin ME, Sauter TC, et al. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. BMJ Open. 2016;6.
9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165.
10. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. PLoS ONE. 2018;13.
11. Eames J, Eisenman A, Schuster RJ. Disagreement between emergency department admission diagnosis and hospital discharge diagnosis: mortality and morbidity. Diagnosis (Berl). 2016;3.

12. Ben-Assuli O, Sagi D, Leshno M, Ironi A, Ziv A. Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. J Biomed Inform. 2015;55.

13. Bernhard M, Raatz C, Zahn P, Merker A, Gries A. Validity of admission diagnoses as process-driving criteria. Influence on length of stay and consultation rate in emergency departments. Anaesthesist. 2013;62.

14. Macaluso CR, McNamara RM. Evaluation and management of acute abdominal pain in the emergency department. Int J Gen Med. 2012;5.

15. Chiu HS, Chan KF, Chung CH, Ma K, Au KW. A Comparison of Emergency Department Admission Diagnoses and Discharge Diagnoses: Retrospective Study. Hong Kong Journal of Emergency Medicine. 2003;10.

16. Kryzauskas M, Danys D, Poskus T, Mikalauskas S, Poskus E, Jotautas V, et al. Is acute appendicitis still misdiagnosed? Open Med (Wars). 2016;11.

17. Johnson T, McNutt R, Odwazny R, Patel D, Baker S. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. J Hosp Med. 2009;4.

18. McWilliams A, Tapp H, Barker J, Dulin M. Cost analysis of the use of emergency departments for primary care services in Charlotte, North Carolina. N C Med J. 2011;72.

19. Mistry B, Stewart De Ramirez S, Kelen G, Schmitz PSK, Balhara KS, Levin S, et al. Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment. Ann Emerg Med. 2018;71.

20. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform. 2016;Suppl 1.

21. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, Bower P. The Effectiveness of Electronic Differential Diagnoses (DDX) Generators: A Systematic Review and Meta-Analysis. PLoS ONE. 2016;11.

22. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users. JMIR Form Res. 2019;3.

23. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis. 2019;14.

24. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med. 2008;23 Suppl 1.

25. El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. BMJ Qual Saf. 2013;22 Suppl 2.

26. Hirsch JA, Leslie-Mazwi TM, Nicola GN, Oklu R, Schoppe KA, Silva E, Manchikanti L. The ICD-10 system: a gift that keeps on taking. J Neurointerv Surg. 2015;7.

27. Hall DE, Arya S, Schmid KK, Blaser C, Carlson MA, Bailey TL, et al. Development and Initial Validation of the Risk Analysis Index for Measuring Frailty in Surgical Populations. JAMA Surg. 2017;152.

28. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994;47.

29. Clavien P-A, Vetter D, Staiger RD, Slankamenac K, Mehra T, Graf R, Puhan MA. The Comprehensive Complication Index (CCI®): Added Value and Clinical Perspectives 3 Years "Down the Line". Ann Surg. 2017;265.

30. Clavien PA, Barkun J, Oliveira ML de, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250.

31. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med. 2012;27.

32. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas. 2018;30.

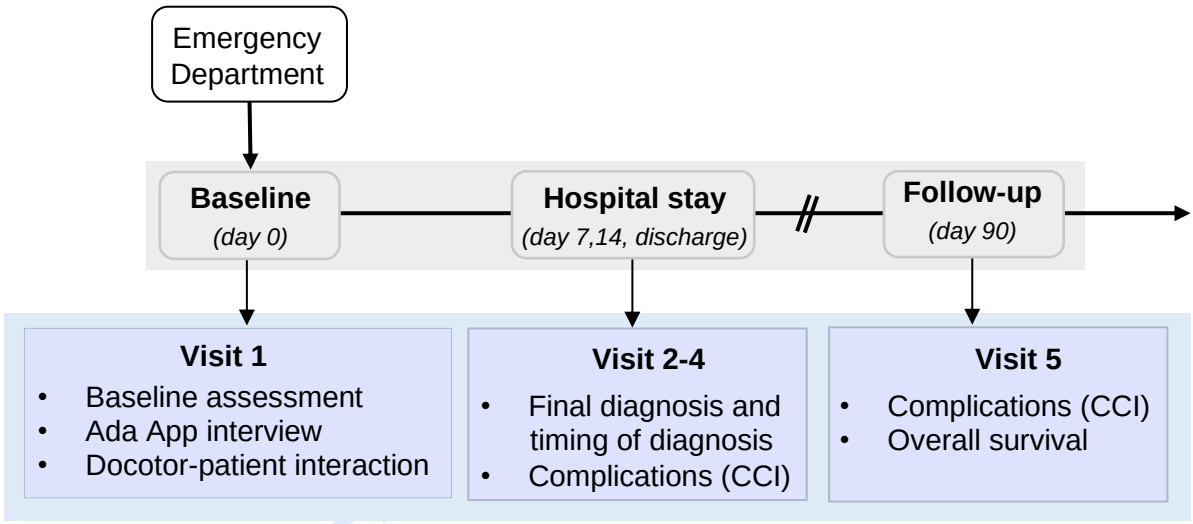
33. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. Ann Emerg Med. 2018;71.

34. Ramnarayan P, Cronje N, Brown R, Negus R, Coode B, Moss P, et al. Validation of a diagnostic reminder system in emergency medicine: a multi-centre study. Emerg Med J. 2007;24.

35. Ekstrom HL, Kharbanda EO, Ballard DW, Vinson DR, Vazquez-Benitez G, Chettipally UK, et al. Development of a Clinical Decision Support System for Pediatric Abdominal Pain in Emergency Department Settings Across Two Health Systems Within the HCSRN. EGEMS (Wash DC) 2019.

36. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25.

37. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA. 2018;320.



Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-207

		Reporting Item	Page Number
Administrative information			
Title	#1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	#2a	Trial identifier and registry name. If not yet registered, name of intended registry	2
Trial registration: data set	#2b	All items from the World Health Organization Trial Registration Data Set	n/a
Protocol version	#3	Date and version identifier	24/03/2020, version 2.0
Funding	#4	Sources and types of financial, material, and other support	15
Roles and responsibilities: contributorship	#5a	Names, affiliations, and roles of protocol contributors	1, 15

Roles and responsibilities: sponsor contact information	#5b	Name and contact information for the trial sponsor	1
Roles and responsibilities: sponsor and funder	#5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	n/a, no sponsors or funders
Roles and responsibilities: committees	#5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	n/a
Introduction			
Background and rationale	#6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	4-5
Background and rationale: choice of comparators	#6b	Explanation for choice of comparators	n/a
Objectives	#7	Specific objectives or hypotheses	2
Trial design	#8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory)	6-9
Methods: Participants, interventions, and outcomes			
Study setting	#9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	6

Eligibility criteria	#10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	6
Interventions: description	#11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	n/a, no interventions planned
Interventions: modifications	#11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease)	n/a
Interventions: adherance	#11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests)	n/a
Interventions: concomitant care	#11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	n/a
Outcomes	#12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	9-10
Participant timeline	#13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	6-8
Sample size	#14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	11-12
Recruitment	#15	Strategies for achieving adequate participant enrolment to reach target sample size	11-12
Methods: Assignment of interventions (for			

controlled trials)			
Allocation: sequence generation	#16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	n/a
Allocation concealment mechanism	#16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	n/a
Allocation: implementation	#16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	n/a
Blinding (masking)	#17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	n/a
Blinding (masking): emergency unblinding	#17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	n/a
Methods: Data collection, management, and analysis			
Data collection plan	#18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	10-11
Data collection plan: retention	#18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	10-11
Data management	#19	Plans for data entry, coding, security, and storage, including	10-11

		any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	
Statistics: outcomes	#20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	11-12
Statistics: additional analyses	#20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	11-12
Statistics: analysis population and missing data	#20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	11-12
Methods: Monitoring			
Data monitoring: formal committee	#21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	n/a
Data monitoring: interim analysis	#21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	11-12
Harms	#22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	10
Auditing	#23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	n/a
Ethics and dissemination			
Research ethics approval	#24	Plans for seeking research ethics committee / institutional review board (REC / IRB) approval	2,12
Protocol amendments	#25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to	2

		relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators)	
Consent or assent	#26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	6,12
Consent or assent: ancillary studies	#26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	n/a
Confidentiality	#27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	10-11
Declaration of interests	#28	Financial and other competing interests for principal investigators for the overall trial and each study site	15
Data access	#29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	15-16
Ancillary and post trial care	#30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	15-16
Dissemination policy: trial results	#31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	11,16
Dissemination policy: authorship	#31b	Authorship eligibility guidelines and any intended use of professional writers	n/a
Dissemination policy: reproducible research	#31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	n/a
Appendices			
Informed consent materials	#32	Model consent form and other related documentation given to participants and authorised surrogates	6,12
Biological specimens	#33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the	n/a

current trial and for future use in ancillary studies, if applicable

Notes:

- 3: 24/03/2020, version 2.0
- 5c: n/a, no sponsors or funders
- 11a: n/a, no interventions planned The SPIRIT checklist is distributed under the terms of the Creative Commons Attribution License CC-BY-ND 3.0. This checklist was completed on 06. June 2020 using https://www.goodreports.org/, a tool made by the EQUATOR Network in collaboration with Penelope.ai

BMJ Open

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041396.R1
Article Type:	Protocol
Date Submitted by the Author:	04-Nov-2020
Complete List of Authors:	Faqar-Uz-Zaman, S. Fatima; Hospital of the Goethe University Frankfurt Surgery Centre, Department for General, Visceral and Transplant Surgery Filmann, Natalie; Institute of Biostatistics and Mathematical Modeling, Goethe-University, Frankfurt/Main, Mahkovic, Dora; Ljubljana Central Medical School von Wagner, Michael; Goethe University Frankfurt Detemble, Charlotte ; Hospital of the Goethe University Frankfurt Surgery Centre Kippke, Ulf; Hospital of the Goethe University Frankfurt Surgery Centre Marschall, Ursula; BARMER Anantharajah, Luxia ; Hospital of the Goethe University Frankfurt Surgery Centre Baumartz, Philipp; Hospital of the Goethe University Frankfurt Surgery Centre Sobotta, Paula ; Hospital of the Goethe University Frankfurt Surgery Centre Bechstein, Wolf; Hospital of the Goethe University Frankfurt Surgery Centre Schnitzbauer, Andreas; Hospital of the Goethe University Frankfurt Surgery Centre
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Emergency medicine, Gastroenterology and hepatology, Diagnostics
Keywords:	ACCIDENT & EMERGENCY MEDICINE, Adult gastroenterology < GASTROENTEROLOGY, Adult surgery < SURGERY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

S. Fatima Faqar-Uz-Zaman[1], Natalie Filmann[2], Dora Mahkovic[3], Michael von Wagner[4], Charlotte Detemble[1], Ulf Kippke[1], Ursula Marschall[5], Luxia Anantharajah[1], Philipp Baumartz[1], Paula Sobotta[1], Wolf O. Bechstein[1] and Andreas A. Schnitzbauer[1]

[1] Department for General, Visceral and Transplant Surgery, Frankfurt University Hospital, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

[2] Institute of Biostatistics and Mathematical Modeling, Goethe-University Frankfurt, Germany

[3] MCL Medical Center Ljubljana, Ljubljana, Slovenia

[4] Executive Department for Medical IT-Systems and Digitalization, Frankfurt University Hospital, Germany

[5] Barmer health insurance, Germany

Corresponding author:

Dr. S. Fatima Faqar-Uz-Zaman

Department for General, Visceral and Transplant Surgery

Frankfurt University Hospital, Goethe-University Frankfurt/Main

Theodor-Stern-Kai 7

60590 Frankfurt am Main

Phone: +49-69-6301-5253

Fax: +49-69-6301-84028

Email: sarafatima.faqar-uz-zaman@kgu.de

Abstract

Introduction: Occurrence of inaccurate or delayed diagnoses is a significant concern in patient care, particularly in emergency medicine, where decision-making is often constrained by high throughput and inaccurate admission diagnoses. Artificial intelligence (AI)-based diagnostic decision support system (DDSS) have been developed to enhance clinical performance by suggesting differential diagnoses to a given case, based on an integrated medical knowledge base and machine learning techniques. The purpose of the study is to evaluate the diagnostic accuracy of Ada®, an app-based diagnostic tool, and the impact on patient outcome.

Methods and analysis: The eRadaR trial is a prospective, double-blinded study with patients presenting to the emergency room (ER) with abdominal pain. At initial contact in the ER, a structured interview will be performed using the Ada-App® and both, patients and attending physicians, will be blinded to the proposed diagnosis lists until trial completion. Throughout the study, clinical data relating to diagnostic findings and types of therapy will be obtained and the follow-up until day 90 will comprise occurrence of complications and overall survival of patients. The primary efficacy of the trial is defined by the percentage of correct diagnoses suggested by Ada® compared to the final discharge diagnosis. Further, accuracy and timing of diagnosis will be compared to decision-making of classical doctor-patient interaction. Secondary objectives are complications, length of hospital stay, and overall survival.

Ethics and dissemination: Ethical approval was received by the independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 including the patient information material and informed consent form. All protocol amendments must be reported to and adapted by the IEC. The results from this study will be submitted to peer-reviewed journals and reported at suitable national and international meetings.

Trial registration: German Clinical Trials Register: DRKS00019098 (registered on 29th May 2020).

Keywords: Artificial intelligence, diagnostic accuracy, emergency room diagnoses, abdominal pain

Article Summary

Strengths and weaknesses of this study

- This is the first prospective study to examine the diagnostic accuracy of an app-based diagnostic tool in an emergency room (ED) and the impact on clinical outcomes.
- The study will be conducted in a real-life setting to investigate the performance in a high stress environment and to provide rational for routine clinical application.
- The double-blinded design will avoid bias regarding research findings.
- The primary limitation of an observational design is that only associations can be described, not causal relationships.

Introduction

Diagnostic errors, comprising inaccurate, delayed, or missed diagnoses, are one of the major challenges in public healthcare [1]. In the recent 'Patient Safety Fact File', the World Health Organization (WHO) outlines ten crucial facts about patient safety [2]. Accordingly, adverse events are among the ten leading causes of death and disability, contributing to approximately 10% of patients harmed during hospitalization. Of note, 10% to 20% of adverse events have been quoted to be particularly related to diagnostic failure, causing more harm to patients than medication or treatment errors [3–5]. Further, false or delayed diagnoses are reported to be the most common reason for medical malpractice litigation [6]. Graber *et al.* estimated that diagnostic failures occurred in 5%-15% of cases, depending on the medical specialty with higher percentages assumed in primary care and emergency medicine [7]. Various reasons have been identified to contribute to false diagnoses. Graber concluded that cognitive slips, primarily resulting from faulty information processing and verification, and misguided situational confidence occur most frequently [8, 9].

This is especially evident in ER settings, which often have to deal with high throughputs, fast decision-making, and incomplete clinical information in a disruptive environment. In particular, ER overcrowding has been identified as a serious threat to patient safety, resulting in poor clinical outcome and a significant increase in mortality [10].

Previous studies have revealed that more than 40% of admission diagnoses at first presentation to the ER are not concordant with the final diagnosis of the patient [11–13]. That means, that throughout the hospital stay, the patient experiences a change in diagnosis based on a variety of additional diagnostics and reevaluation of initial assumptions, finally leading to the correct diagnosis. In particular, approximately 30% of patients with abdominal pain, being one of the leading causes for visiting the ER, exhibit a discrepancy in diagnosis [14, 15]. In particular, misdiagnosis rate of acute appendicitis, the most frequent reason for acute abdominal pain, has largely remained unchanged over time and is still associated with a high ratio of negative appendectomies [16]. Inaccurate diagnosing in ERs has been shown to be further associated with increased length of hospital stay, rate of consultations, healthcare cost, and risk for mortality and morbidity, contributing to a serious concern to patient safety [11, 13, 17, 18]. Thus, a high degree of diagnostic accuracy can lead to an

improvement in quality of patient care. Correct admission diagnoses are crucial for a reliable triage and process management and critically influence the initial evaluation in that ER and subsequent clinical course of the patient [19].

Digital technologies and artificial intelligence (AI)-based methods have recently emerged as impressively powerful tools to empower physicians in clinical decision making and improve healthcare quality. More specifically, diagnostic decision support systems (DDSS) have demonstrated to facilitate assessment of clinical data input by using an extensive medical knowledge base [20, 21]. One version of DDSS is Ada®, an app-based AI-machine learning system that incorporates patients' symptoms and other findings into its knowledge base and intelligent technology to deliver effective healthcare [22, 23]. Based on an algorithmic pathway and driven by chief complaints, the app-based system generates a set of differential diagnoses for a given clinical case. Several studies have reported that DDSS have the potential to increase diagnostic performance, obtaining an accuracy rate of 70-96% [24, 25]. In particular, a retrospective study of rare diseases has demonstrated that Ada suggests accurate diagnoses earlier than clinical diagnoses in more than half of all cases [23].

However, application of the Ada app has not been investigated in a real-life setting, particularly in ERs, which has to deal with a high stress environment and heavy time constraints. This app-based method may be a valuable companion in triaging patients and support clinicians in making decisions more accurate and sooner by simultaneously reducing risk for medical errors. Therefore, in the present study, we aim to evaluate the diagnostic ability of Ada in ER settings and examine the impact on timing of diagnosis.

Methods and analysis

The eRadaR-trial is designed as a prospective, double-blinded, observational study evaluating the diagnostic accuracy of the Ada-App® in the ER of the Department of General, Visceral and Transplant Surgery of the Frankfurt University Hospital, Germany. This trial is registered as DRKS00019098 in the German Clinical Trials Register and the trial protocol is written in accordance with the current Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT 2013). The SPIRIT checklist is given in Additional file 1.

Study population and eligibility criteria

All patients presenting to the ER with abdominal pain will form the study population and be screened for trial eligibility. Patients, who will be immediately discharged from the ED on the same day and patients, who will be admitted to the hospital after presenting to the ER will be both included in the study and followed up. Inclusion criteria comprise: (a.) adults aged ≥ 18 years, (b.) patients presenting with abdominal pain to the ER, and (c.) patients willing to participate and able to provide written informed consent. The criteria of exclusion are: (a.) intubated patients, (b.) unstable patients or (c.) patients with severe injuries requiring immediate medical treatment, (d.) patients unwilling or incapable of providing informed consent. Eligible patients are asked for their participation in the trial and written informed consent will be obtained from themselves. All reasons for exclusion of patients will be recorded in the trial screening log and analyzed accordingly.

Description of study visits and assessment schedule

Eligible patients will be interviewed by the study team with the Ada-App® based on an algorithmic pathway of questions relating to the symptoms. The Ada-App® will only obtain data about patient demographics, patient history, and information about current complaints. Patient's name and date of birth will be pseudonymized using an individual identification code, as described in the section 'data management and data safety'. Throughout the study, the patient, the study team, and the physician treating the patient will be blinded regarding the list of proposed diagnoses by the app. The patient will subsequently be diagnosed by classical doctor-patient interaction and decision-making. The clinical course of the patient will be followed until day 90 after initial contact in the ER. Detailed information about outline of the study and assessment schedule are displayed in Table1 and Figure1.

Table 1 Schedule of study visits and assessments of the eRadaR study

	Baseline	**Hospital stay***	**Discharge**	**90-days FU**
Visits	V1 (Day 0)	V2 and 3 (Day 7, 14)	V4	V5 (Day 90)
Informed consent	X			
Eligibility criteria	X			
Demographic data: a) Charlson Comorbidity Index b) RAI-C score	X X			X
Ada diagnosis list	X			
ICD-10 diagnoses	X		X	X
Symptoms	X			
Diagnostics**		X	X	
Therapy and OPS-code		X	X	
Rate of consultations			X	
Complications (CCI)		X	X	X
Length of hospital stay			X	
Overall survival				X

* Visit 2 or 3 are left out, if the patient is discharged before
** Diagnostics include:
a) Routine blood samples (C-reactive protein, White blood cells, Hemoglobin, Platelets, Sodium, Potassium, Creatinine, Albumin, Bilirubin, INR)
b) Instrumental diagnostics (Ultrasound, Chest/Abdominal CT/MRI, ECG, Endoscopy)
RAI-C = Risk Analysis C score; ICD = International Classification of Diseases; OPS = Operations and Procedures;
CCI = Comprehensive Complication Index, FU = Follow-up, V = visit

Patient presenting to the ED (Visit 1)

After enrollment in the trial, a structured interview with the Ada-App® will be conducted and bassline data will be assessed including demographic data according to the Carlson Comorbidity Index and the Risk Analysis Index-C score (RAI-C score), the patients' symptoms and ICD-10 diagnoses list [26–28]. Participants are then diagnosed and treated according to the standard of care by the attending physician of the ER. As this is a double-blinded study to patients and treating physicians, Ada-App®

diagnoses lists will be randomly allocated to a study-ID and then manually transferred into the electronic case report forms (eCRF). The trial personnel will be blinded until the end of the study to avoid bias regarding subsequent diagnoses and treatment of the patient, except of the interim analysis, which is mentioned in the section of statistical analysis.

Hospital stay (Visit 2, day 7)

This visit is performed on day 7, after the patient is admitted to the hospital. Data about diagnostics and therapies are assessed comprising laboratory results (i.e. C-reactive protein, white blood cells, platelets, hemoglobin, bilirubin, creatinine, sodium and potassium, albumin, INR), computer-assisted diagnostics (i.e. ultrasound, chest/abdominal CT/MRI, ECG, endoscopy), type of therapy (conservative, interventional or surgery), OPS-code of therapies and complications according to the Comprehensive Complication Index (CCI) together with the date of occurrence [29]. If the patient has not been admitted to the hospital or is discharged before day 7, visit 2 is left out.

Hospital stay (Visit 3, day 14)

Visit 3 is performed on day 14 after patient's admission and assessment schedule is equivalent to visit 2. If the patient has not been admitted to the hospital or is discharged before day 14, visit 3 is left out.

Discharge (Visit 4)

At discharge, data including the final ICD-10 diagnosis and the timing of diagnosis will be recorded to subsequently analyze the accuracy and the timing of the Ada-App® compared to the classical doctor-patient-encounter. Further data items include diagnostics (laboratory, instrumental), OPS-codes and type of therapies, complications according to CCI, length of hospital stay, overall health cost, rate of consultation.

Follow-up (Visit 5)

The follow-up will be performed as a structured telephone interview or in person on day 90 and will encompass following data items: demographic data according to the RAI-C score, complication assessment according to CCI and overall survival.

Interventions

As this is an observational, double-blinded, prospective study, no experimental or control interventions are conducted.

Endpoints

Primary endpoint

The primary endpoint of this study is to evaluate the diagnostic accuracy of the Ada-App® by comparing the decision-making of the classical doctor-patient interaction with the diagnoses proposed by the app-based algorithm.

Secondary endpoint

Secondary endpoints of this study consist of the following: timing of final discharge diagnosis and time to treatment during hospital stay, comparing accurate diagnoses with discharge diagnoses as descriptive assessments, the occurrence of complications according to the CCI, total length of stay in hospital from initial contact in the ER until discharge, patient morbidity and mortality at day 90, overall health cost analysis and consultation rate. Further endpoints are displayed in the description of assessment schedule (Table1).

Measurement methods

For data capture, following measurement methods will be used:

1. Primary outcome measurement will be performed using the Ada-App® which will deliver a set of differential diagnoses to a given clinical case [23]. Based on an algorithmic questionnaire and machine learning technologies, the Ada chatbot assesses symptoms of the patient, similar to the anamnestic techniques and clinical reasoning of physicians. Patients' data are integrated into an extensive knowledge base, which has been specifically designed by medical doctors by incorporating validated disease models and comprehensive medical literature. Then, differential diagnoses are generated and ranked in order considering two features: the probability, based on epidemiologic data, and the best match between the diagnosis and the given symptoms. Through AI-based methods and multiple feedback loops, the Ada® knowledge base grows after each interaction and diagnostic ability improves continuously.
2. The occurrence of complications as secondary outcomes will be evaluated and analyzed according to the Comprehensive Complication Index (CCI) [29]. The CCI represents the standard assessment of postoperative morbidity and comprises all

complications occurring during a patient's course based on the Clavien-Dindo classification (CDC). Compared to the CDC, which ranks complications based on the severity of the therapeutic consequence and grades them in 5 levels, the CCI uses a formula to integrate all complications, ranging them from 0 ('no complication') to 100 ('death') [30]. This advanced approach enables comparison of patients harboring more than one complication and takes more subtle differences into consideration.

3. For assessment of comorbid diseases and frailty-associated risk in a surgical population, we will use the Charlson Comorbidity Index and the RAI-C score.

Risk-Benefit assessment

This is an observational, non-interventional study and does not comprise any specific risk for the patient, as data obtained with the app are not used in the ER standard of care. Therefore, there is no special need for additional safety management. A delay in the diagnosis and treatment of patients presenting to the ER is not expected, as the app-based interview will not require more than 10 minutes and will exclusively be performed in the waiting zone of the ER by the study team. Baseline assessment (during visit 1) will directly be conducted after patient has been registered at the ER and given informed consent. Besides that, unstable patients requiring immediate medical care are excluded from the study beforehand.

Data management and data safety

The investigators will design and produce electronic case report forms (eCRF) for protocol-required data collection. All information will be entered into these eCRFs by authorized and trained members of the study team and systematically checked for accuracy and completeness. Staff members with responsibilities for data collection or those, having access to the database will be enrolled in a delegation log. Patients' data collected during the trial will be recorded in pseudonymized form by solely using individual identification codes.

For data assessment using the Ada-App®, a specified iPad will be provided, which will be registered at the Frankfurt University Hospital and will be exclusively used for the purpose of this trial. Clinical data will be documented pseudonymously by using a combination of a random number from 1 to 450 and the patient's year of birth. Participants are then asked to answer the questionnaire of the Ada-App® preferably by

themselves or otherwise assisted by the study team. The diagnoses will be manually transferred into the eCRF of the related patient after trial completion and unblinding.

All trial data obtained will be integrated into a statistical analysis software and analyzed by the Institute of Biostatistics and Mathematical Modeling Frankfurt.

Publication policy

The results of this trial will be submitted for publication in a peer-reviewed journal in a summarized anonymized manner. The study is scientifically supported by the Barmer health insurance company. Barmer will act as a scientific advisor regarding the conduct of the study, will be involved in the process of interpreting the data and in the publication and public distribution process of the study after trial completion. However, there will be no raw data sharing or financial support from the institution.

Statistical analysis

Interim analysis

One formal unblinded interim analysis of the trial data is planned to be performed after enrollment of about 200 patients to evaluate the diagnostic accuracy of the Ada-App® with 90-days follow-up information. Statistical analysis will be performed by the responsible study biometrician using a significance level of alpha = 0.001 and a subsequent report will be written. These results will be discussed with the investigators and the study team in a staff meeting and the continuation of the trial will be considered.

Sample size calculation and study duration

The assumptions that were made, was that more than 30% of the admission diagnoses are not consistent with the final discharge diagnosis and hypothesized that the Ada-App® will increase the diagnostic accuracy from 70% to a rate of 85%. Providing a power of 90% and a two-sided significance level of alpha 5%, a target sample size of N = 405 patients has to be recruited to detect the targeted effect. With an estimated dropout rate of 10%, we plan to recruit N = 450 patients in this trial. Furthermore, we expect the width of the confidence intervals for the diagnostic accuracy to be 0.1 at maximum (0.09 with an estimated diagnostic accuracy of 0.7, 0.07 with an estimated diagnostic accuracy of 0.85).

This trial is anticipated to start in September 2020 and the duration of patient's participation is 3 months including follow-up. To achieve the required sample size of patients, trial completion is expected to be in 12 months (August 2020).

Ethical and legal aspects, consent

The eRadaR trial will be conducted in accordance with the Declaration of Helsinki and the international conference of harmonization good clinical practice (ICH-GCP) guidelines. After a patient has been identified to meet eligibility criteria, the patient will be informed about the aim, outline and individual risk of the study and informed consent will be given. After a sufficient period, the patient can then sign informed consent and will receive a signed copy.

Patient and Public involvement

Patients were not involved in the development of the research question or study design. They will however be involved in visit 1 and will be interviewed by the study team using the Ada-App®. Further, the follow-up (visit 5) will be performed as a telephone interview or in person with the patients for data assessment.

Discussion

Diagnostic errors have been identified as a serious threat to patient safety, leading to preventable adverse events, particularly in ERs with a disruptive environment. AI-based tools and algorithms have the potential to substantially reduce diagnostic failures, achieving high rates of diagnostic accuracy, which rivals the capability of clinicians.

A previous study provides an overview of the main types of existing tools, which are classified into categories related to the targeted step of diagnostic processing [25]. Over the past few decades, a number of computerized DDSS have been developed, exhibiting promising diagnostic efficacy. Bond *et al.* evaluated four current DDSS using clinical cases from the *New England Journal of Medicine*, demonstrating that Isabel and Dxplain achieve the strongest performance [31]. Compared to former programs, second-generation DDSS are far more powerful, providing more accurate suggestions with increasing complexity, while concomitantly requiring less time for diagnosing [21, 24]. This is primarily essential in an era of ER crowding, where fast and accurate triaging is necessary to prioritize critically ill patients and to optimize resource allocation [8]. Stewart *et al.* recently summarized various fields of AI application becoming relevant in emergency medicine, including imaging, decision-making, and outcome prediction [32]. In terms of triaging, a machine learning-based tool efficiently predicts critical patient outcome, equivalent to the classically used Emergency Severity Index [33]. In a prospective, multi-center study, the DDSS Isabel achieved high accuracy in diagnosing patients presenting to the ER, suggesting the final discharge diagnosis in 95% of cases [34]. Another clinical decision support system has been evaluated in patients presenting with acute abdominal pain aiming to identify high-risk patients for acute appendicitis [35]. Based on automated methods and an integrated risk calculator, patient data was assessed from the electronic health record (EHR) and management strategies suggested according to the risk level. Incorporation into EHR represents one of the most recent advances in the development of DDSS using 'natural language processing' techniques, which matches entered clinical data with the underlying knowledge base [36]. This might facilitate assessment of larger volumes of data, save more time, and might increase acceptance of DDSS in clinical workflow.

However, in most of these trials using clinical support systems, impact on patient outcome, or on healthcare costs were not assessed. Although diagnoses suggested

by DDSS mostly contained the correct diagnosis and achieved high level of users' satisfaction, relevance, and specificity of extensive lists were low [21, 24, 31]. Long lists may lead to distraction or to unnecessary diagnostic with increased risk for iatrogenic injuries and costs. In general, despite the given potential efficacy of DDSS, widespread acceptance for implementation of DDSS into the routine clinical practice is evolving scarcely [37]. Studies focusing on AI-based diagnostic tools are generally designed heterogeneously and are often of poor quality, making it difficult to recommend widespread evidence-based clinical application [21, 25]. While most of the current trials demonstrated high diagnostic accuracy in retrospective and simulated cases, only few studies evaluated their performance in real clinical settings, particularly in high stress environments like ERs. Thus, further validations in prospective studies are required to investigate the diagnostic efficiency and utility of DDSS and their impact on routine clinical decision-making and patient outcome.

Trial status

Ethical approval for this trial was granted by an independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 and anticipated trial start date is September 2020.

Abbreviations

AI, Artificial Intelligence; App, Application; CCI, Comprehensive Complication Index; CDC, Clavien-Dindo classification; DRKS, Deutsches Register Klinische Studien (German Clinical Trials Register); eCRF, electronic case report form; DDSS, Diagnostic Decision Support System; ER, Emergency Room; EHR, Electronic Health Record; FU, Follow-Up; ICD-10, International Classification of Diseases; ICH-GCP, International Conference of Harmonisation, Good Clinical Practice; IEC, Independent ethics committee; OPS, Operations and Procedures; RAI-C, Risk Analysis Index-C score.

Acknowledgments

We would like to thank Rene Wendel, Doreen Giannico, Dr. Johannes Masseli, Prof. Stefan Zeuzem, and the whole administrative and medical team of the central emergency room of the Frankfurt University Hospital for their support.

We further are grateful for our established scientific support with the Barmer health care insurance company.

Funding

There is no funding for this trial from third parties or grants. This is an investigator-initiated trial.

Competing interest

No competing interests were detected for the specific trial.

Author Contributions

SFF has written the manuscript, was involved in writing the protocol, and is the clinical lead surgeon for the trial.

NF performed the sample size calculation for the trial and was involved in drafting the protocol and the manuscript.

MvW supported the trial as Chief Medical Informatics Officer (CMIO) of the hospital, and was involved in drafting the protocol and the manuscript

CD and DM were involved in creating the idea and drafting the protocol and the manuscript and prepared the CRF logistics.

UK is the quality manager of the surgical department and supervised implementation of the project in the ER.

UM is the BARMER health insurance representative and is an external advisor to the trial. She drafted the protocol and the manuscript.

LA, PB, PS are medical students triaging the patients and putting up logistics in the ER. They all were involved in creating the idea and shaping the project.

WOB gave valuable input into the project, supports it majorly as chair of the department, and creates a culture for innovative projects. He further drafted the protocol and manuscript.

AAS had the idea for the project, leads the trial group and links all parties involved. He is the responsible principal investigator for the trial.

Availability of data and materials

Data and material will be available after an adequate research proposal has been made and reviewed by the responsible persons at the University. The requests will be forwarded to our Ethics committee and the requesting party will be responsible to get

permission to do posthoc analysis and will be also liable for any costs arising from the requests.

Consent for publication

All authors read and consented to the publication of the current version of this manuscript.

References

1. Balla J, Heneghan C, Goyder C, Thompson M. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. BMJ Open 2012.
2. World Health Organization. Patient Safety and Risk Management Service Delivery and Safety. Accessed Sept 2019. https://www.who.int/features/factfiles/patient_safety/en/.
3. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington (DC); 2000.
4. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998;169.
5. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324.
6. Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, et al. Claims, errors, and compensation payments in medical malpractice litigation. N Engl J Med. 2006;354.
7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121.
8. Hautz SC, Schuler L, Kämmer JE, Schauber SK, Ricklin ME, Sauter TC, et al. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. BMJ Open. 2016;6.
9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165.
10. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. PLoS ONE. 2018;13.
11. Eames J, Eisenman A, Schuster RJ. Disagreement between emergency department admission diagnosis and hospital discharge diagnosis: mortality and morbidity. Diagnosis (Berl). 2016;3.

12. Ben-Assuli O, Sagi D, Leshno M, Ironi A, Ziv A. Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. J Biomed Inform. 2015;55.

13. Bernhard M, Raatz C, Zahn P, Merker A, Gries A. Validity of admission diagnoses as process-driving criteria. Influence on length of stay and consultation rate in emergency departments. Anaesthesist. 2013;62.

14. Macaluso CR, McNamara RM. Evaluation and management of acute abdominal pain in the emergency department. Int J Gen Med. 2012;5.

15. Chiu HS, Chan KF, Chung CH, Ma K, Au KW. A Comparison of Emergency Department Admission Diagnoses and Discharge Diagnoses: Retrospective Study. Hong Kong Journal of Emergency Medicine. 2003;10.

16. Kryzauskas M, Danys D, Poskus T, Mikalauskas S, Poskus E, Jotautas V, et al. Is acute appendicitis still misdiagnosed? Open Med (Wars). 2016;11.

17. Johnson T, McNutt R, Odwazny R, Patel D, Baker S. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. J Hosp Med. 2009;4.

18. McWilliams A, Tapp H, Barker J, Dulin M. Cost analysis of the use of emergency departments for primary care services in Charlotte, North Carolina. N C Med J. 2011;72.

19. Mistry B, Stewart De Ramirez S, Kelen G, Schmitz PSK, Balhara KS, Levin S, et al. Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment. Ann Emerg Med. 2018;71.

20. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform. 2016;Suppl 1.

21. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, Bower P. The Effectiveness of Electronic Differential Diagnoses (DDX) Generators: A Systematic Review and Meta-Analysis. PLoS ONE. 2016;11.

22. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users. JMIR Form Res. 2019;3.

23. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis. 2019;14.

24. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med. 2008;23 Suppl 1.

25. El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. BMJ Qual Saf. 2013;22 Suppl 2.

26. Hirsch JA, Leslie-Mazwi TM, Nicola GN, Oklu R, Schoppe KA, Silva E, Manchikanti L. The ICD-10 system: a gift that keeps on taking. J Neurointerv Surg. 2015;7.

27. Hall DE, Arya S, Schmid KK, Blaser C, Carlson MA, Bailey TL, et al. Development and Initial Validation of the Risk Analysis Index for Measuring Frailty in Surgical Populations. JAMA Surg. 2017;152.

28. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994;47.

29. Clavien P-A, Vetter D, Staiger RD, Slankamenac K, Mehra T, Graf R, Puhan MA. The Comprehensive Complication Index (CCI®): Added Value and Clinical Perspectives 3 Years "Down the Line". Ann Surg. 2017;265.

30. Clavien PA, Barkun J, Oliveira ML de, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250.

31. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med. 2012;27.

32. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas. 2018;30.

33. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. Ann Emerg Med. 2018;71.

34. Ramnarayan P, Cronje N, Brown R, Negus R, Coode B, Moss P, et al. Validation of a diagnostic reminder system in emergency medicine: a multi-centre study. Emerg Med J. 2007;24.

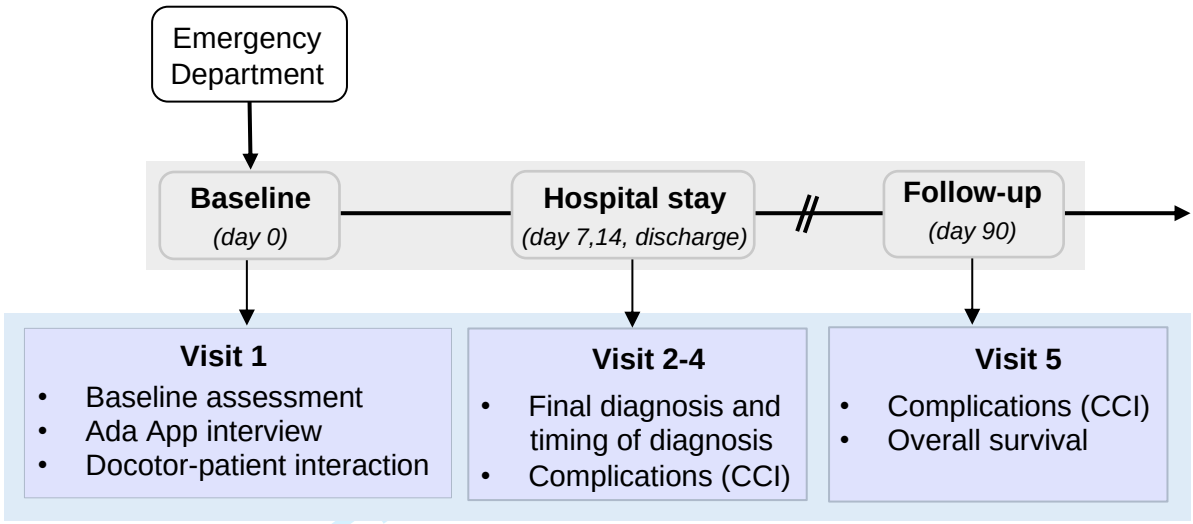
35. Ekstrom HL, Kharbanda EO, Ballard DW, Vinson DR, Vazquez-Benitez G, Chettipally UK, et al. Development of a Clinical Decision Support System for Pediatric Abdominal Pain in Emergency Department Settings Across Two Health Systems Within the HCSRN. EGEMS (Wash DC) 2019.

36. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25.

37. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA. 2018;320.

Figure legend

Figure 1: Study flow chart of the eRadaR study. CCI = Comprehensive complication index.



Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-207

		Reporting Item	Page Number
Administrative information			
Title	#1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	#2a	Trial identifier and registry name. If not yet registered, name of intended registry	2
Trial registration: data set	#2b	All items from the World Health Organization Trial Registration Data Set	n/a
Protocol version	#3	Date and version identifier	24/03/2020, version 2.0
Funding	#4	Sources and types of financial, material, and other support	15
Roles and	#5a	Names, affiliations, and roles of protocol contributors	1, 15

responsibilities: contributorship			
Roles and responsibilities: sponsor contact information	#5b	Name and contact information for the trial sponsor	1
Roles and responsibilities: sponsor and funder	#5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	n/a, no sponsors or funders
Roles and responsibilities: committees	#5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	n/a
Introduction			
Background and rationale	#6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	4-5
Background and rationale: choice of comparators	#6b	Explanation for choice of comparators	n/a
Objectives	#7	Specific objectives or hypotheses	2
Trial design	#8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory)	6-9
Methods: Participants, interventions, and outcomes			

Study setting	#9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	6
Eligibility criteria	#10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	6
Interventions: description	#11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	n/a, no interventions planned
Interventions: modifications	#11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease)	n/a
Interventions: adherance	#11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests)	n/a
Interventions: concomitant care	#11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	n/a
Outcomes	#12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	9-10
Participant timeline	#13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	6-8
Sample size	#14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions	11-12

		supporting any sample size calculations	
Recruitment	#15	Strategies for achieving adequate participant enrolment to reach target sample size	11-12
Methods: Assignment of interventions (for controlled trials)			
Allocation: sequence generation	#16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	n/a
Allocation concealment mechanism	#16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	n/a
Allocation: implementation	#16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	n/a
Blinding (masking)	#17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	n/a
Blinding (masking): emergency unblinding	#17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	n/a
Methods: Data collection, management, and analysis			
Data collection plan	#18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate	10-11

		measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	
Data collection plan: retention	#18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	10-11
Data management	#19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	10-11
Statistics: outcomes	#20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	11-12
Statistics: additional analyses	#20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	11-12
Statistics: analysis population and missing data	#20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	11-12
Methods: Monitoring			
Data monitoring: formal committee	#21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	n/a
Data monitoring:	#21b	Description of any interim analyses and stopping	11-12

interim analysis		guidelines, including who will have access to these interim results and make the final decision to terminate the trial	
Harms	#22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	10
Auditing	#23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	n/a
Ethics and dissemination			
Research ethics approval	#24	Plans for seeking research ethics committee / institutional review board (REC / IRB) approval	2,12
Protocol amendments	#25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators)	2
Consent or assent	#26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	6,12
Consent or assent: ancillary studies	#26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	n/a
Confidentiality	#27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	10-11
Declaration of interests	#28	Financial and other competing interests for principal investigators for the overall trial and each study site	15
Data access	#29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	15-16

Ancillary and post trial care	#30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	15-16
Dissemination policy: trial results	#31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	11,16
Dissemination policy: authorship	#31b	Authorship eligibility guidelines and any intended use of professional writers	n/a
Dissemination policy: reproducible research	#31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	n/a
Appendices			
Informed consent materials	#32	Model consent form and other related documentation given to participants and authorised surrogates	6,12
Biological specimens	#33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	n/a

Notes:

- 3: 24/03/2020, version 2.0
- 5c: n/a, no sponsors or funders
- 11a: n/a, no interventions planned The SPIRIT checklist is distributed under the terms of the Creative Commons Attribution License CC-BY-ND 3.0. This checklist was completed on 06. June 2020 using https://www.goodreports.org/, a tool made by the EQUATOR Network in collaboration with Penelope.ai

Note from the Editors: Instructions for reviewers of study protocols

Since launching in 2011, BMJ Open has published study protocols for planned or ongoing research studies. If data collection is complete, we will not consider the manuscript.

Publishing study protocols enables researchers and funding bodies to stay up to date in their fields by providing exposure to research activity that may not otherwise be widely publicised. This can help prevent unnecessary duplication of work and will hopefully enable collaboration. Publishing protocols in full also makes available more information than is currently required by trial registries and increases transparency, making it easier for others (editors, reviewers and readers) to see and understand any deviations from the protocol that occur during the conduct of the study.

The scientific integrity and the credibility of the study data depend substantially on the study design and methodology, which is why the study protocol requires a thorough peer-review.

BMJ Open will consider for publication protocols for any study design, including observational studies and systematic reviews.

Some things to keep in mind when reviewing the study protocol:

- Protocol papers should report planned or ongoing studies. The dates of the study should be included in the manuscript.
- Unfortunately we are unable to customize the reviewer report form for study protocols. As such, some of the items (i.e., those pertaining to results) on the form should be scores as Not Applicable (N/A).
- While some baseline data can be presented, there should be no results or conclusions present in the study protocol.
- For studies that are ongoing, it is generally the case that very few changes can be made to the methodology. As such, requests for revisions are generally clarifications for the rationale or details relating to the methods. If there is a major flaw in the study that would prevent a sound interpretation of the data, we would expect the study protocol to be rejected.

BMJ Open

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041396.R2
Article Type:	Protocol
Date Submitted by the Author:	07-Dec-2020
Complete List of Authors:	Faqar-Uz-Zaman, S. Fatima; Hospital of the Goethe University Frankfurt Surgery Centre, Department for General, Visceral and Transplant Surgery Filmann, Natalie; Institute of Biostatistics and Mathematical Modeling, Goethe-University, Frankfurt/Main, Mahkovic, Dora; Ljubljana Central Medical School von Wagner, Michael; Goethe University Frankfurt Detemble, Charlotte ; Hospital of the Goethe University Frankfurt Surgery Centre Kippke, Ulf; Hospital of the Goethe University Frankfurt Surgery Centre Marschall, Ursula; BARMER Anantharajah, Luxia ; Hospital of the Goethe University Frankfurt Surgery Centre Baumartz, Philipp; Hospital of the Goethe University Frankfurt Surgery Centre Sobotta, Paula ; Hospital of the Goethe University Frankfurt Surgery Centre Bechstein, Wolf; Hospital of the Goethe University Frankfurt Surgery Centre Schnitzbauer, Andreas; Hospital of the Goethe University Frankfurt Surgery Centre
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Emergency medicine, Gastroenterology and hepatology, Diagnostics
Keywords:	ACCIDENT & EMERGENCY MEDICINE, Adult gastroenterology < GASTROENTEROLOGY, Adult surgery < SURGERY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR-trial

S. Fatima Faqar-Uz-Zaman[1], Natalie Filmann[2], Dora Mahkovic[3], Michael von Wagner[4], Charlotte Detemble[1], Ulf Kippke[1], Ursula Marschall[5], Luxia Anantharajah[1], Philipp Baumartz[1], Paula Sobotta[1], Wolf O. Bechstein[1] and Andreas A. Schnitzbauer[1]

[1] Department for General, Visceral and Transplant Surgery, Frankfurt University Hospital, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

[2] Institute of Biostatistics and Mathematical Modeling, Goethe-University Frankfurt, Germany

[3] MCL Medical Center Ljubljana, Ljubljana, Slovenia

[4] Executive Department for Medical IT-Systems and Digitalization, Frankfurt University Hospital, Germany

[5] Barmer health insurance, Germany

Corresponding author:

Dr. S. Fatima Faqar-Uz-Zaman

Department for General, Visceral and Transplant Surgery

Frankfurt University Hospital, Goethe-University Frankfurt/Main

Theodor-Stern-Kai 7

60590 Frankfurt am Main

Phone: +49-69-6301-5253

Fax: +49-69-6301-84028

Email: SaraFatima.Faqar-Uz-Zaman@kgu.de

Abstract

Introduction: Occurrence of inaccurate or delayed diagnoses is a significant concern in patient care, particularly in emergency medicine, where decision-making is often constrained by high throughput and inaccurate admission diagnoses. Artificial intelligence (AI)-based diagnostic decision support system (DDSS) have been developed to enhance clinical performance by suggesting differential diagnoses to a given case, based on an integrated medical knowledge base and machine learning techniques. The purpose of the study is to evaluate the diagnostic accuracy of Ada®, an app-based diagnostic tool, and the impact on patient outcome.

Methods and analysis: The eRadaR trial is a prospective, double-blinded study with patients presenting to the emergency room (ER) with abdominal pain. At initial contact in the ER, a structured interview will be performed using the Ada-App® and both, patients and attending physicians, will be blinded to the proposed diagnosis lists until trial completion. Throughout the study, clinical data relating to diagnostic findings and types of therapy will be obtained and the follow-up until day 90 will comprise occurrence of complications and overall survival of patients. The primary efficacy of the trial is defined by the percentage of correct diagnoses suggested by Ada® compared to the final discharge diagnosis. Further, accuracy and timing of diagnosis will be compared to decision-making of classical doctor-patient interaction. Secondary objectives are complications, length of hospital stay, and overall survival.

Ethics and dissemination: Ethical approval was received by the independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 including the patient information material and informed consent form. All protocol amendments must be reported to and adapted by the IEC. The results from this study will be submitted to peer-reviewed journals and reported at suitable national and international meetings.

Trial registration: German Clinical Trials Register: DRKS00019098 (registered on 29th May 2020).

Keywords: Artificial intelligence, diagnostic accuracy, emergency room diagnoses, abdominal pain

Article Summary

Strengths and weaknesses of this study

- This is the first prospective study to examine the diagnostic accuracy of an app-based diagnostic tool in an emergency room (ED) and the impact on clinical outcomes.
- The study will be conducted in a real-life setting to investigate the performance in a high stress environment and to provide rational for routine clinical application.
- The double-blinded design will avoid bias regarding research findings.
- The primary limitation of an observational design is that only associations can be described, not causal relationships.

Introduction

Diagnostic errors, comprising inaccurate, delayed, or missed diagnoses, are one of the major challenges in public healthcare [1]. In the recent 'Patient Safety Fact File', the World Health Organization (WHO) outlines ten crucial facts about patient safety [2]. Accordingly, adverse events are among the ten leading causes of death and disability, contributing to approximately 10% of patients harmed during hospitalization. Of note, 10% to 20% of adverse events have been quoted to be particularly related to diagnostic failure, causing more harm to patients than medication or treatment errors [3–5]. Further, false or delayed diagnoses are reported to be the most common reason for medical malpractice litigation [6]. Graber *et al.* estimated that diagnostic failures occurred in 5%-15% of cases, depending on the medical specialty with higher percentages assumed in primary care and emergency medicine [7]. Various reasons have been identified to contribute to false diagnoses. Graber concluded that cognitive slips, primarily resulting from faulty information processing and verification, and misguided situational confidence occur most frequently [8, 9].

This is especially evident in ER settings, which often have to deal with high throughputs, fast decision-making, and incomplete clinical information in a disruptive environment. In particular, ER overcrowding has been identified as a serious threat to patient safety, resulting in poor clinical outcome and a significant increase in mortality [10].

Previous studies have revealed that more than 40% of admission diagnoses at first presentation to the ER are not concordant with the final diagnosis of the patient [11–13]. That means, that throughout the hospital stay, the patient experiences a change in diagnosis based on a variety of additional diagnostics and reevaluation of initial assumptions, finally leading to the correct diagnosis. In particular, approximately 30% of patients with abdominal pain, being one of the leading causes for visiting the ER, exhibit a discrepancy in diagnosis [14, 15]. In particular, misdiagnosis rate of acute appendicitis, the most frequent reason for acute abdominal pain, has largely remained unchanged over time and is still associated with a high ratio of negative appendectomies [16]. Inaccurate diagnosing in ERs has been shown to be further associated with increased length of hospital stay, rate of consultations, healthcare cost, and risk for mortality and morbidity, contributing to a serious concern to patient safety [11, 13, 17, 18]. Thus, a high degree of diagnostic accuracy can lead to an

improvement in quality of patient care. Correct admission diagnoses are crucial for a reliable triage and process management and critically influence the initial evaluation in that ER and subsequent clinical course of the patient [19].

Digital technologies and artificial intelligence (AI)-based methods have recently emerged as impressively powerful tools to empower physicians in clinical decision making and improve healthcare quality. More specifically, diagnostic decision support systems (DDSS) have demonstrated to facilitate assessment of clinical data input by using an extensive medical knowledge base [20, 21]. One version of DDSS is Ada®, an app-based AI-machine learning system that incorporates patients' symptoms and other findings into its knowledge base and intelligent technology to deliver effective healthcare [22, 23]. Based on an algorithmic pathway and driven by chief complaints, the app-based system generates a set of differential diagnoses for a given clinical case. Several studies have reported that DDSS have the potential to increase diagnostic performance, obtaining an accuracy rate of 70-96% [24, 25]. In particular, a retrospective study of rare diseases has demonstrated that Ada suggests accurate diagnoses earlier than clinical diagnoses in more than half of all cases [23].

However, application of the Ada app has not been investigated in a real-life setting, particularly in ERs, which has to deal with a high stress environment and heavy time constraints. This app-based method may be a valuable companion in triaging patients and support clinicians in making decisions more accurate and sooner by simultaneously reducing risk for medical errors. Therefore, in the present study, we aim to evaluate the diagnostic ability of Ada in ER settings and examine the impact on timing of diagnosis.

Methods and analysis

The eRadaR-trial is designed as a prospective, double-blinded, observational study evaluating the diagnostic accuracy of the Ada-App® in the ER of the Department of General, Visceral and Transplant Surgery of the Frankfurt University Hospital, Germany. This trial is registered as DRKS00019098 in the German Clinical Trials Register and the trial protocol is written in accordance with the current Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT 2013). The SPIRIT checklist is given in Additional file 1.

The Ada-App® specifications and rationale to use the software

The Ada-App® is a class I medicinal product certified in accordance with the DIN ISO 13485. Ada is a free-downloadable certified medicinal product and has been validated in different studies by the marketing authorization holder and developer team. It has shown a higher accuracy (73%) in comparison to other apps (38%) when compared to the correctness of symptom checking. The App was superior to other apps when the hitlist of the 5 most probable diagnoses were compared (84% vs. 51%) [22, 26–32].

The evidence shows that the algorithm is superior to other solutions on the market, it has been validated by the company, and the data were the basis for the certification as a medicinal product class I (CE-mark in accordance with DIN ISO 13485), supporting our rationale to test the potentially most beneficial and promising software on the market.

Study population and eligibility criteria

All patients presenting to the ER with abdominal pain will form the study population and be screened for trial eligibility. Notably, patients presenting with abdominal pain as part of multiple chief complaints (e.g. chest pain and abdominal pain) will also be included in the study. Moreover, patients, who will be immediately discharged from the ED on the same day and patients, who will be admitted to the hospital after presenting to the ER will be both included in the study and followed up in an intention-to-treat fashion. Inclusion criteria comprise: (a.) adults aged ≥ 18 years, (b.) patients presenting with abdominal pain to the ER, and (c.) patients willing to participate and able to provide written informed consent. The criteria of exclusion are: (a.) intubated patients, (b.) unstable patients or (c.) patients with severe injuries requiring immediate medical treatment, (d.) patients unwilling or incapable of providing informed consent. Eligible

patients are asked for their participation in the trial and written informed consent will be obtained from themselves. All reasons for exclusion of patients will be recorded in the trial screening log and analyzed accordingly.

Description of study visits and assessment schedule

Eligible patients will be interviewed by the study team with the Ada-App® based on an algorithmic pathway of questions relating to the symptoms. The Ada-App® will only obtain data about patient demographics, patient history, and information about current complaints. Patient's name and date of birth will be pseudonymized using an individual identification code, as described in the section 'data management and data safety'. Throughout the study, the patient, the study team, and the physician treating the patient will be blinded regarding the list of proposed diagnoses by the app. The patient will subsequently be diagnosed by classical doctor-patient interaction and decision-making. The clinical course of the patient will be followed until day 90 after initial contact in the ER. Detailed information about outline of the study and assessment schedule are displayed in Table1 and Figure1.

Table 1 Schedule of study visits and assessments of the eRadaR study

	Baseline	Hospital stay*	Discharge	90-days FU
Visits	V1 (Day 0)	V2 and 3 (Day 7, 14)	V4	V5 (Day 90)
Informed consent	X			
Eligibility criteria	X			
Demographic data:	X			X
a) Charlson Comorbidity Index	X			
b) RAI-C score				
Ada diagnosis list	X			
ICD-10 diagnoses	X		X	X
Symptoms	X			
Diagnostics**		X	X	
Therapy and OPS-code		X	X	
Rate of consultations			X	
Complications (CCI)		X	X	X
Length of hospital stay			X	
Overall survival				X

* Visit 2 or 3 are left out, if the patient is discharged before
** Diagnostics include:
a) Routine blood samples (C-reactive protein, White blood cells, Hemoglobin, Platelets, Sodium, Potassium, Creatinine, Albumin, Bilirubin, INR)
b) Instrumental diagnostics (Ultrasound, Chest/Abdominal CT/MRI, ECG, Endoscopy)
RAI-C = Risk Analysis C score; ICD = International Classification of Diseases; OPS = Operations and Procedures;
CCI = Comprehensive Complication Index, FU = Follow-up, V = visit

Patient presenting to the ED (Visit 1)

After enrollment in the trial, a structured interview with the Ada-App® will be conducted and bassline data will be assessed including demographic data according to the Carlson Comorbidity Index and the Risk Analysis Index-C score (RAI-C score), the patients' symptoms and ICD-10 diagnoses list [33–35]. Participants are then diagnosed and treated according to the standard of care by the attending physician of the ER. As this is a double-blinded study to patients and treating physicians, Ada-App®

diagnoses lists will be randomly allocated to a study-ID and then manually transferred into the electronic case report forms (eCRF). The trial personnel will be blinded until the end of the study to avoid bias regarding subsequent diagnoses and treatment of the patient, except of the interim analysis, which is mentioned in the section of statistical analysis.

Hospital stay (Visit 2, day 7)

This visit is performed on day 7, after the patient is admitted to the hospital. Data about diagnostics and therapies are assessed comprising laboratory results (i.e. C-reactive protein, white blood cells, platelets, hemoglobin, bilirubin, creatinine, sodium and potassium, albumin, INR), computer-assisted diagnostics (i.e. ultrasound, chest/abdominal CT/MRI, ECG, endoscopy), type of therapy (conservative, interventional or surgery), OPS-code of therapies and complications according to the Comprehensive Complication Index (CCI) together with the date of occurrence [36]. If the patient has not been admitted to the hospital or is discharged before day 7, visit 2 is left out.

Hospital stay (Visit 3, day 14)

Visit 3 is performed on day 14 after patient's admission and assessment schedule is equivalent to visit 2. If the patient has not been admitted to the hospital or is discharged before day 14, visit 3 is left out.

Discharge (Visit 4)

At discharge, data including the final ICD-10 diagnosis and the timing of diagnosis will be recorded to subsequently analyze the accuracy and the timing of the Ada-App® compared to the classical doctor-patient-encounter. Further data items include diagnostics (laboratory, instrumental), OPS-codes and type of therapies, complications according to CCI, length of hospital stay, overall health cost, rate of consultation.

Follow-up (Visit 5)

The follow-up will be performed as a structured telephone interview or in person on day 90 and will encompass following data items: demographic data according to the RAI-C score, complication assessment according to CCI and overall survival.

Interventions

As this is an observational, double-blinded, prospective study, no experimental or control interventions are conducted.

Endpoints

Primary endpoint

The primary endpoint of this study is to evaluate the diagnostic accuracy of the Ada-App® by comparing the decision-making of the classical doctor-patient interaction with the diagnoses proposed by the app-based algorithm.

Secondary endpoint

Secondary endpoints of this study consist of the following: timing of final discharge diagnosis and time to treatment during hospital stay, comparing accurate diagnoses with discharge diagnoses as descriptive assessments, the occurrence of complications according to the CCI, total length of stay in hospital from initial contact in the ER until discharge, patient morbidity and mortality at day 90, overall health cost analysis and consultation rate. Further endpoints are displayed in the description of assessment schedule (Table1).

Measurement methods

For data capture, following measurement methods will be used:

1. Primary outcome measurement will be performed using the Ada-App® which will deliver a set of differential diagnoses to a given clinical case [23]. Based on an algorithmic questionnaire and machine learning technologies, the Ada chatbot assesses symptoms of the patient, similar to the anamnestic techniques and clinical reasoning of physicians. Patients' data are integrated into an extensive knowledge base, which has been specifically designed by medical doctors by incorporating validated disease models and comprehensive medical literature. Then, differential diagnoses are generated and ranked in order considering two features: the probability, based on epidemiologic data, and the best match between the diagnosis and the given symptoms. Through AI-based methods and multiple feedback loops, the Ada® knowledge base grows after each interaction and diagnostic ability improves continuously.
2. The occurrence of complications as secondary outcomes will be evaluated and analyzed according to the Comprehensive Complication Index (CCI) [36]. The CCI represents the standard assessment of postoperative morbidity and comprises all

complications occurring during a patient's course based on the Clavien-Dindo classification (CDC). Compared to the CDC, which ranks complications based on the severity of the therapeutic consequence and grades them in 5 levels, the CCI uses a formula to integrate all complications, ranging them from 0 ('no complication') to 100 ('death') [37]. This advanced approach enables comparison of patients harboring more than one complication and takes more subtle differences into consideration.

3. For assessment of comorbid diseases and frailty-associated risk in a surgical population, we will use the Charlson Comorbidity Index and the RAI-C score.

Risk-Benefit assessment

This is an observational, non-interventional study and does not comprise any specific risk for the patient, as data obtained with the app are not used in the ER standard of care. Therefore, there is no special need for additional safety management. A delay in the diagnosis and treatment of patients presenting to the ER is not expected, as the app-based interview will not require more than 10 minutes and will exclusively be performed in the waiting zone of the ER by the study team. Baseline assessment (during visit 1) will directly be conducted after patient has been registered at the ER and given informed consent. Besides that, unstable patients requiring immediate medical care are excluded from the study beforehand.

Data management and data safety

The investigators will design and produce electronic case report forms (eCRF) for protocol-required data collection. All information will be entered into these eCRFs by authorized and trained members of the study team and systematically checked for accuracy and completeness. Staff members with responsibilities for data collection or those, having access to the database will be enrolled in a delegation log. Patients' data collected during the trial will be recorded in pseudonymized form by solely using individual identification codes.

For data assessment using the Ada-App®, a specified iPad will be provided, which will be registered at the Frankfurt University Hospital and will be exclusively used for the purpose of this trial. Clinical data will be documented pseudonymously by using a combination of a random number from 1 to 450 and the patient's year of birth. Participants are then asked to answer the questionnaire of the Ada-App® preferably by

themselves or otherwise assisted by the study team. The diagnoses will be manually transferred into the eCRF of the related patient after trial completion and unblinding.

All trial data obtained will be integrated into a statistical analysis software and analyzed by the Institute of Biostatistics and Mathematical Modeling Frankfurt.

Ethics and dissemination

The eRadaR trial will be conducted in accordance with the Declaration of Helsinki and the international conference of harmonization good clinical practice (ICH-GCP) guidelines. After a patient has been identified to meet eligibility criteria, the patient will be informed about the aim, outline and individual risk of the study and informed consent will be given. After a sufficient period, the patient can then sign informed consent and will receive a signed copy.

The results of this trial will be submitted for publication in a peer-reviewed journal in a summarized anonymized manner. The study is scientifically supported by the Barmer health insurance company. Barmer will act as a scientific advisor regarding the conduct of the study, will be involved in the process of interpreting the data and in the publication and public distribution process of the study after trial completion. However, there will be no raw data sharing or financial support from the institution.

Statistical analysis

Interim analysis

One formal unblinded interim analysis of the trial data is planned to be performed after enrollment of about 200 patients to evaluate the diagnostic accuracy of the Ada-App® with 90-days follow-up information. Statistical analysis will be performed by the responsible study biometrician using a significance level of alpha = 0.001 and a subsequent report will be written. These results will be discussed with the investigators and the study team in a staff meeting and the continuation of the trial will be considered.

Sample size calculation and study duration

The assumptions that were made, was that more than 30% of the admission diagnoses are not consistent with the final discharge diagnosis and hypothesized that the Ada-App® will increase the diagnostic accuracy from 70% to a rate of 85%. Providing a power of 90% and a two-sided significance level of alpha 5%, a target sample size of N = 405 patients has to be recruited to detect the targeted effect. With an estimated

dropout rate of 10%, we plan to recruit *N* = 450 patients in this trial. Furthermore, we expect the width of the confidence intervals for the diagnostic accuracy to be 0.1 at maximum (0.09 with an estimated diagnostic accuracy of 0.7, 0.07 with an estimated diagnostic accuracy of 0.85).

This trial is anticipated to start in September 2020 and the duration of patient's participation is 3 months including follow-up. To achieve the required sample size of patients, trial completion is expected to be in 12 months (August 2020).

Patient and Public involvement

Patients were not involved in the development of the research question or study design. They will however be involved in visit 1 and will be interviewed by the study team using the Ada-App®. Further, the follow-up (visit 5) will be performed as a telephone interview or in person with the patients for data assessment.

Discussion

Diagnostic errors have been identified as a serious threat to patient safety, leading to preventable adverse events, particularly in ERs with a disruptive environment. AI-based tools and algorithms have the potential to substantially reduce diagnostic failures, achieving high rates of diagnostic accuracy, which rivals the capability of clinicians.

A previous study provides an overview of the main types of existing tools, which are classified into categories related to the targeted step of diagnostic processing [25]. Over the past few decades, a number of computerized DDSS have been developed, exhibiting promising diagnostic efficacy. Bond *et al.* evaluated four current DDSS using clinical cases from the *New England Journal of Medicine*, demonstrating that Isabel and Dxplain achieve the strongest performance [38]. Compared to former programs, second-generation DDSS are far more powerful, providing more accurate suggestions with increasing complexity, while concomitantly requiring less time for diagnosing [21, 24]. This is primarily essential in an era of ER crowding, where fast and accurate triaging is necessary to prioritize critically ill patients and to optimize resource allocation [8]. Stewart *et al.* recently summarized various fields of AI application becoming relevant in emergency medicine, including imaging, decision-making, and outcome prediction [39]. In terms of triaging, a machine learning-based tool efficiently predicts critical patient outcome, equivalent to the classically used Emergency Severity Index [40]. In a prospective, multi-center study, the DDSS Isabel achieved high accuracy in diagnosing patients presenting to the ER, suggesting the final discharge diagnosis in 95% of cases [41]. Another clinical decision support system has been evaluated in patients presenting with acute abdominal pain aiming to identify high-risk patients for acute appendicitis [42]. Based on automated methods and an integrated risk calculator, patient data was assessed from the electronic health record (EHR) and management strategies suggested according to the risk level. Incorporation into EHR represents one of the most recent advances in the development of DDSS using 'natural language processing' techniques, which matches entered clinical data with the underlying knowledge base [43]. This might facilitate assessment of larger volumes of data, save more time, and might increase acceptance of DDSS in clinical workflow.

However, in most of these trials using clinical support systems, impact on patient outcome, or on healthcare costs were not assessed. Although diagnoses suggested

by DDSS mostly contained the correct diagnosis and achieved high level of users' satisfaction, relevance, and specificity of extensive lists were low [20, 25, 38]. Long lists may lead to distraction or to unnecessary diagnostic with increased risk for iatrogenic injuries and costs. In general, despite the given potential efficacy of DDSS, widespread acceptance for implementation of DDSS into the routine clinical practice is evolving scarcely [44]. Studies focusing on AI-based diagnostic tools are generally designed heterogeneously and are often of poor quality, making it difficult to recommend widespread evidence-based clinical application [21, 25]. While most of the current trials demonstrated high diagnostic accuracy in retrospective and simulated cases, only few studies evaluated their performance in real clinical settings, particularly in high stress environments like ERs. Thus, further validations in prospective studies are required to investigate the diagnostic efficiency and utility of DDSS and their impact on routine clinical decision-making and patient outcome.

Trial status

Ethical approval for this trial was granted by an independent ethics committee (IEC) of the Goethe-University Frankfurt on 9th April 2020 and anticipated trial start date is September 2020.

Abbreviations

AI, Artificial Intelligence; App, Application; CCI, Comprehensive Complication Index; CDC, Clavien-Dindo classification; DRKS, Deutsches Register Klinische Studien (German Clinical Trials Register); eCRF, electronic case report form; DDSS, Diagnostic Decision Support System; ER, Emergency Room; EHR, Electronic Health Record; FU, Follow-Up; ICD-10, International Classification of Diseases; ICH-GCP, International Conference of Harmonisation, Good Clinical Practice; IEC, Independent ethics committee; OPS, Operations and Procedures; RAI-C, Risk Analysis Index-C score.

Acknowledgments

We would like to thank Rene Wendel, Doreen Giannico, Dr. Johannes Masseli, Prof. Stefan Zeuzem, and the whole administrative and medical team of the central emergency room of the Frankfurt University Hospital for their support.

We further are grateful for our established scientific support with the Barmer health care insurance company.

Funding

There is no funding for this trial from third parties or grants. This is an investigator-initiated trial.

Competing interest

No competing interests were detected for the specific trial.

Author Contributions

SFF has written the manuscript, was involved in writing the protocol, and is the clinical lead surgeon for the trial.

NF performed the sample size calculation for the trial and was involved in drafting the protocol and the manuscript.

MvW supported the trial as Chief Medical Informatics Officer (CMIO) of the hospital, and was involved in drafting the protocol and the manuscript

CD and DM were involved in creating the idea and drafting the protocol and the manuscript and prepared the CRF logistics.

UK is the quality manager of the surgical department and supervised implementation of the project in the ER.

UM is the BARMER health insurance representative and is an external advisor to the trial. She drafted the protocol and the manuscript.

LA, PB, PS are medical students triaging the patients and putting up logistics in the ER. They all were involved in creating the idea and shaping the project.

WOB gave valuable input into the project, supports it majorly as chair of the department, and creates a culture for innovative projects. He further drafted the protocol and manuscript.

AAS had the idea for the project, leads the trial group and links all parties involved. He is the responsible principal investigator for the trial.

Availability of data and materials

Data and material will be available after an adequate research proposal has been made and reviewed by the responsible persons at the University. The requests will be forwarded to our Ethics committee and the requesting party will be responsible to get

permission to do posthoc analysis and will be also liable for any costs arising from the requests.

Consent for publication

All authors read and consented to the publication of the current version of this manuscript.

References

1. Balla J, Heneghan C, Goyder C, Thompson M. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. BMJ Open 2012.
2. World Health Organization. Patient Safety and Risk Management Service Delivery and Safety. Accessed Sept 2019. https://www.who.int/features/factfiles/patient_safety/en/.
3. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324.
4. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington (DC); 2000.
5. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998;169.
6. Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, et al. Claims, errors, and compensation payments in medical malpractice litigation. N Engl J Med. 2006;354.
7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121.
8. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165.
9. Hautz SC, Schuler L, Kämmer JE, Schauber SK, Ricklin ME, Sauter TC, et al. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. BMJ Open. 2016;6.
10. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. PLoS ONE. 2018;13.
11. Bernhard M, Raatz C, Zahn P, Merker A, Gries A. Validity of admission diagnoses as process-driving criteria. Influence on length of stay and consultation rate in emergency departments. Anaesthesist. 2013;62.
12. Ben-Assuli O, Sagi D, Leshno M, Ironi A, Ziv A. Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. J Biomed Inform. 2015;55.

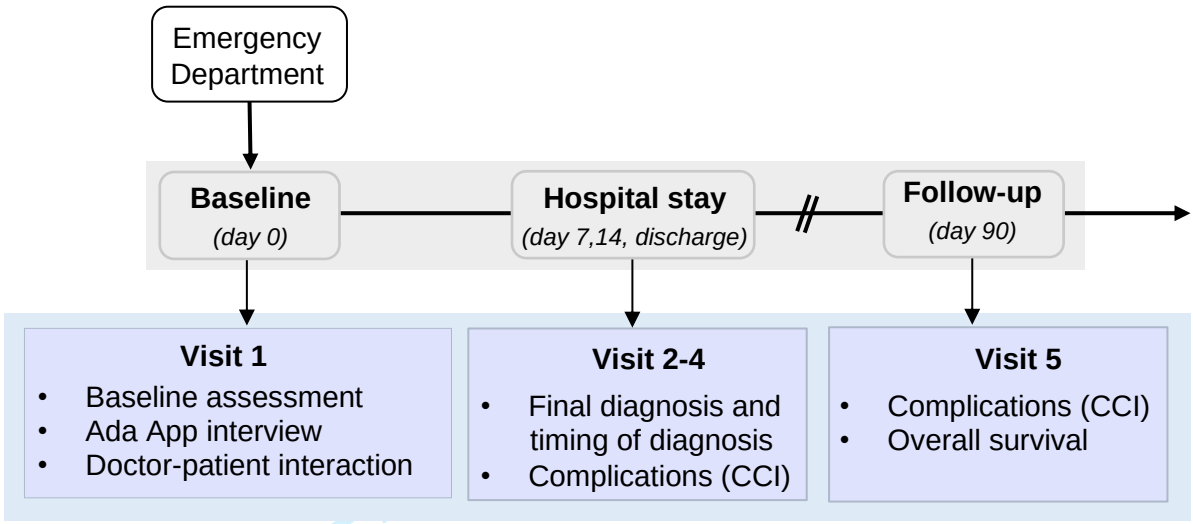
13. Eames J, Eisenman A, Schuster RJ. Disagreement between emergency department admission diagnosis and hospital discharge diagnosis: mortality and morbidity. Diagnosis (Berl). 2016;3.
14. Chiu HS, Chan KF, Chung CH, Ma K, Au KW. A Comparison of Emergency Department Admission Diagnoses and Discharge Diagnoses: Retrospective Study. Hong Kong Journal of Emergency Medicine. 2003;10.
15. Macaluso CR, McNamara RM. Evaluation and management of acute abdominal pain in the emergency department. Int J Gen Med. 2012;5.
16. Kryzauskas M, Danys D, Poskus T, Mikalauskas S, Poskus E, Jotautas V, et al. Is acute appendicitis still misdiagnosed? Open Med (Wars). 2016;11.
17. McWilliams A, Tapp H, Barker J, Dulin M. Cost analysis of the use of emergency departments for primary care services in Charlotte, North Carolina. N C Med J. 2011;72.
18. Johnson T, McNutt R, Odwazny R, Patel D, Baker S. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. J Hosp Med. 2009;4.
19. Mistry B, Stewart De Ramirez S, Kelen G, Schmitz PSK, Balhara KS, Levin S, et al. Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment. Ann Emerg Med. 2018;71.
20. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, Bower P. The Effectiveness of Electronic Differential Diagnoses (DDX) Generators: A Systematic Review and Meta-Analysis. PLoS ONE. 2016;11.
21. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform. 2016;Suppl 1.
22. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet journal of rare diseases 2019.
23. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users. JMIR Form Res. 2019;3.
24. El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. BMJ Qual Saf. 2013;22 Suppl 2.

25. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med. 2008;23 Suppl 1.
26. Montazeri M, Multmeier J, Novorol C, Upadhyay S, Wicks P, Gilbert S. The potential for digital patient symptom recording through symptom assessment applications to optimize patient flow and reduce waiting times in Urgent Care Centers: a simulation study; 2020.
27. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors. 2020;7.
28. Mehl A, Bergey F, Cawley C, Gilsdorf A. Syndromic surveillance insights from a symptom assessment app before and during COVID-19 measures in Germany and the United Kingdom: results from repeated cross-sectional analyses; 2020.
29. Knitza J, Callhoff J, Chehab G, Hueber A, Kiltz U, Kleyer A, et al. Positionspapier der Kommission Digitale Rheumatologie der Deutschen Gesellschaft für Rheumatologie e. V.: Aufgaben, Ziele und Perspektiven für eine moderne Rheumatologie. Z Rheumatol. 2020;79.
30. Hirsch MC, Ronicke S, Krusche M, Wagner AD. Rare diseases 2030: how augmented AI will support diagnosis and treatment of rare diseases in the future. Annals of the rheumatic diseases 2020.
31. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. Original research: How accurate are digital symptom assessment apps for suggesting conditions and urgency advice?: a clinical vignettes comparison to GPs; 2020.
32. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation; 2020.
33. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994;47.
34. Hall DE, Arya S, Schmid KK, Blaser C, Carlson MA, Bailey TL, et al. Development and Initial Validation of the Risk Analysis Index for Measuring Frailty in Surgical Populations. JAMA Surg. 2017;152.
35. Hirsch JA, Leslie-Mazwi TM, Nicola GN, Oklu R, Schoppe KA, Silva E, Manchikanti L. The ICD-10 system: a gift that keeps on taking. J Neurointerv Surg. 2015;7.

36. Clavien P-A, Vetter D, Staiger RD, Slankamenac K, Mehra T, Graf R, Puhan MA. The Comprehensive Complication Index (CCI®): Added Value and Clinical Perspectives 3 Years "Down the Line". Ann Surg. 2017;265.
37. Clavien PA, Barkun J, Oliveira ML de, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250.
38. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med. 2012;27.
39. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas. 2018;30.
40. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. Ann Emerg Med. 2018;71.
41. Ramnarayan P, Cronje N, Brown R, Negus R, Coode B, Moss P, et al. Validation of a diagnostic reminder system in emergency medicine: a multi-centre study. Emerg Med J. 2007;24.
42. Ekstrom HL, Kharbanda EO, Ballard DW, Vinson DR, Vazquez-Benitez G, Chettipally UK, et al. Development of a Clinical Decision Support System for Pediatric Abdominal Pain in Emergency Department Settings Across Two Health Systems Within the HCSRN. EGEMS (Wash DC) 2019.
43. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25.
44. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA. 2018;320.

Figure legend

Figure 1: Study flow chart of the eRadaR study. CCI = Comprehensive complication index.



Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-207

		Reporting Item	Page Number
Administrative information			
Title	#1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	#2a	Trial identifier and registry name. If not yet registered, name of intended registry	2
Trial registration: data set	#2b	All items from the World Health Organization Trial Registration Data Set	n/a
Protocol version	#3	Date and version identifier	24/03/2020, version 2.0
Funding	#4	Sources and types of financial, material, and other support	15
Roles and	#5a	Names, affiliations, and roles of protocol contributors	1, 15

responsibilities: contributorship			
Roles and responsibilities: sponsor contact information	#5b	Name and contact information for the trial sponsor	1
Roles and responsibilities: sponsor and funder	#5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	n/a, no sponsors or funders
Roles and responsibilities: committees	#5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	n/a
Introduction			
Background and rationale	#6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	4-5
Background and rationale: choice of comparators	#6b	Explanation for choice of comparators	n/a
Objectives	#7	Specific objectives or hypotheses	2
Trial design	#8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory)	6-9
Methods: Participants, interventions, and outcomes			

Study setting	#9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	6
Eligibility criteria	#10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	6
Interventions: description	#11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	n/a, no interventions planned
Interventions: modifications	#11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease)	n/a
Interventions: adherance	#11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests)	n/a
Interventions: concomitant care	#11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	n/a
Outcomes	#12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	9-10
Participant timeline	#13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	6-8
Sample size	#14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions	11-12

		supporting any sample size calculations	
Recruitment	#15	Strategies for achieving adequate participant enrolment to reach target sample size	11-12
Methods: Assignment of interventions (for controlled trials)			
Allocation: sequence generation	#16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	n/a
Allocation concealment mechanism	#16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	n/a
Allocation: implementation	#16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	n/a
Blinding (masking)	#17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	n/a
Blinding (masking): emergency unblinding	#17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	n/a
Methods: Data collection, management, and analysis			
Data collection plan	#18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate	10-11

		measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	
Data collection plan: retention	#18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	10-11
Data management	#19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	10-11
Statistics: outcomes	#20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	11-12
Statistics: additional analyses	#20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	11-12
Statistics: analysis population and missing data	#20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	11-12
Methods: Monitoring			
Data monitoring: formal committee	#21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	n/a
Data monitoring:	#21b	Description of any interim analyses and stopping	11-12

interim analysis		guidelines, including who will have access to these interim results and make the final decision to terminate the trial	
Harms	#22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	10
Auditing	#23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	n/a
Ethics and dissemination			
Research ethics approval	#24	Plans for seeking research ethics committee / institutional review board (REC / IRB) approval	2,12
Protocol amendments	#25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators)	2
Consent or assent	#26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	6,12
Consent or assent: ancillary studies	#26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	n/a
Confidentiality	#27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	10-11
Declaration of interests	#28	Financial and other competing interests for principal investigators for the overall trial and each study site	15
Data access	#29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	15-16

Ancillary and post trial care	#30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	15-16
Dissemination policy: trial results	#31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	11,16
Dissemination policy: authorship	#31b	Authorship eligibility guidelines and any intended use of professional writers	n/a
Dissemination policy: reproducible research	#31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	n/a
Appendices			
Informed consent materials	#32	Model consent form and other related documentation given to participants and authorised surrogates	6,12
Biological specimens	#33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	n/a

Notes:

- 3: 24/03/2020, version 2.0
- 5c: n/a, no sponsors or funders
- 11a: n/a, no interventions planned The SPIRIT checklist is distributed under the terms of the Creative Commons Attribution License CC-BY-ND 3.0. This checklist was completed on 06. June 2020 using https://www.goodreports.org/, a tool made by the EQUATOR Network in collaboration with Penelope.ai