



BMJ Open Study protocol for a pilot prospective, observational study investigating the condition suggestion and urgency advice accuracy of a symptom assessment app in sub-Saharan Africa: the AFYA-‘Health’ Study

Elizabeth Millen,¹ Nahya Salim,² Hila Azadzoy,¹ Mustafa Miraji Bane,² Lisa O'Donnell,¹ Marcel Schmude ,¹ Philipp Bode,¹ Ewelina Tuerk,¹ Ria Vaidya,¹ Stephen Henry Gilbert ^{1,3}

To cite: Millen E, Salim N, Azadzoy H, *et al.* Study protocol for a pilot prospective, observational study investigating the condition suggestion and urgency advice accuracy of a symptom assessment app in sub-Saharan Africa: the AFYA-‘Health’ Study. *BMJ Open* 2022;**12**:e055915. doi:10.1136/bmjopen-2021-055915

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-055915>).

EM and NS contributed equally.

Received 28 July 2021
Accepted 14 March 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Stephen Henry Gilbert; science@ada.com

ABSTRACT

Introduction Due to a global shortage of healthcare workers, there is a lack of basic healthcare for 4 billion people worldwide, particularly affecting low-income and middle-income countries. The utilisation of AI-based healthcare tools such as symptom assessment applications (SAAs) has the potential to reduce the burden on healthcare systems. The purpose of the AFYA Study (AI-based Assessment of health sYmptoms in Tanzania) is to evaluate the accuracy of the condition suggestions and urgency advice provided by a user on a Swahili language Ada SAA.

Methods and analysis This study is designed as an observational prospective clinical study. The setting is a waiting room of a Tanzanian district hospital. It will include patients entering the outpatient clinic with various conditions and age groups, including children and adolescents. Patients will be asked to use the SAA before proceeding to usual care. After usual care, they will have a consultation with a study-provided physician. Patients and healthcare practitioners will be blinded to the SAA's results. An expert panel will compare the Ada SAA's condition suggestions and urgency advice to usual care and study provided differential diagnoses and triage. The primary outcome measures are the accuracy and comprehensiveness of the Ada SAA evaluated against the gold standard differential diagnoses.

Ethics and dissemination Ethical approval was received by the ethics committee (EC) of Muhimbili University of Health and Allied Sciences with an approval number MUHAS-REC-09-2019-044 and the National Institute for Medical Research, NIMR/HQ/R.8c/Vol. I/922. All amendments to the protocol are reported and adapted on the basis of the requirements of the EC. The results from this study will be submitted to peer-reviewed journals, local and international stakeholders, and will be communicated in editorials/articles by Ada Health.

Trial registration number NCT04958577.

Strengths and limitations of this study

- This is the first prospective study to evaluate the condition suggestion and urgency advice accuracy of a symptom assessment application (SAA) in a low-income and middle-income country.
- The study will be conducted in a real-life setting in a busy primary healthcare facility setting in Africa, addressing a wide range of conditions, including for children and adolescents.
- Having a study-provided physician diagnosis list, in addition to the usual care physician diagnosis list, complemented by expert panel case review, ensures high confidence in the gold standard diagnosis.
- The study is designed to be a pilot study in order to explore the feasibility of a larger clinical investigation assessment which would assess the overall condition suggestion accuracy and comprehensiveness of the SAA for common and uncommon conditions and the general clinic waiting room setting allows comparison to clinical diagnosis making the observational design a strength and not a limitation to this goal.
- The study is a pilot to assist in the design of a larger confirmatory study and the number of patients enrolled reflects this.

INTRODUCTION

Mobile app-based symptom assessment applications (SAAs) are personal health companions for layperson users, which are designed to help people understand and manage their health. They function by asking the user questions regarding their symptoms and their medical history. They provide a symptom assessment which includes suggestions of relevant conditions, advice on urgency and on appropriate next steps in care.¹ Advanced AI-based SAAs



use approaches to dynamically prioritise the order of the questions asked, hence efficiently gathering the optimal information about the user's condition and prioritising the next question asked. Globally, SAA usage, alongside general health-based internet searching, is increasing year-on-year.^{2 3} Users turn to SAAs to better understand the causes of their symptoms, to decide whether to seek care, and to decide on the acuity of care sought; they also perceive that SAAs provide a more personalised assessment of symptoms than is available through internet search engines.⁴⁻⁶ There have been calls, including in a 2019 systematic review, for more studies to assess SAA use, accuracy of condition suggestion, urgency advice and health economic value,^{1 7} and some of these aspects have been addressed in recent studies.^{4-6 8-10} However, there is a need for further research to more completely evaluate SAAs in specific use cases and specific geographies.

An important area of SAA potential is in low- and middle-income countries (LMICs), including those in sub-Saharan Africa.¹¹ There is a global shortage of up to 7.2 million health workers that has contributed to a lack of basic healthcare for an estimated 4 billion people.^{12 13} Although SAAs do not replace healthcare workers, they do have the potential to increase the provision of actionable healthcare information to users in LMICs, thereby empowering individuals to make timely and better-informed decisions about their health.^{11 14 15} It has been shown in high-income countries that access to SAAs can reduce the burden on the healthcare system, by directing people only to seek attention from healthcare facilities when necessary, and in directing them to the most appropriate level of care.^{6 16} Some AI-based tools have already shown potential in improving care provision and efficiency of healthcare utilisation in LMICs and there have already been promising developments in the use of AI-based tools to support care delivery, through supporting healthcare workers.^{11 15} In the medium term, SAAs may have a role in upskilling health workers and in supporting doctors' decision-making. However, before SAA use can be extended to new LMIC use cases, it is important to have specific clinical evaluations to demonstrate the safety and performance of the underlying medical reasoning technologies for LMIC settings, particularly as most health-related AI solutions have been developed in and for high-resource settings, in languages frequently used in these settings, and on the basis of medical data with bias towards these settings.

Although the medical intelligence (meaning the combination of its reasoning engine and medical knowledge) of the SAA to be evaluated in this study has been validated against a set of several thousand internal test cases, which comprise diseases from different medical specialties and include both common and rare diseases, there is also a need to validate the tool in clinical studies. The SAA is currently being evaluated in studies in Europe and in the USA; however the current study will be the first in sub-Saharan Africa. This study will assess an SAA version highly similar to the on-market tool, using the same underlying medical intelligence. The SAA tested

will differ only in that it is a tablet-computer-optimised version, and that it will not provide the health assessment report to the patients or to the physicians, following the approach of Moreno Barriga, et al.¹⁷ This is to avoid bias, and, as the study is designed to be observational, it must therefore have no effect on usual patient care. Additionally to avoid bias, the patient will use the SAA prior to their consultation with the usual care health practitioner.

The results of this study will define the number of participants needed for a larger study through a power calculation; this later study will be able to determine the accuracy, comprehensiveness of the SAA condition-suggestion advice, and urgency-advice level for which this pilot study can only do on a small scale. The following hypothesis will be investigated in the AFYA (AI-based Assessment of health SYMptoms in TANZania) Study: the accuracy and comprehensiveness of the SAA condition-suggestion advice and urgency-advice is appropriate, when compared with that of the usual care physician diagnosis. Appropriateness of advice will consider the point of use of the SAA in the patient's medical journey, which is prior to medical confirmatory tests, such as blood tests and medical imaging. Absolute quantitative targets of medical accuracy have not been specified for testing this hypothesis, but it is expected that the levels of accuracy would be similar to those obtained in the published literature for high-income countries.^{8 17 18}

METHODS AND ANALYSIS

This AFYA Trial is a prospective, observational study conducted at the waiting room of the Mbagala Rangi Tatu Hospital, Dar es Salaam, Tanzania. The trial protocol was developed in accordance with the current Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT).¹⁹ The SPIRIT checklist is available in the supplemental files as online supplemental file 1.

The Ada-App specifications and rationale to use the software

The accuracy and comprehensiveness of the SAA used in this study have been validated in several studies, and it has a higher accuracy (73%) in comparison to other apps (38%) in condition suggestion.⁸ To date, there have been no studies on SAAs thus far in LMICs, creating a need for data to determine its usefulness in these settings.

The SAA evaluated in this study is a European Union (EU) regulatory-approved medical device Conformité Européenne (CE)-marked with study-related modifications, which was developed by Ada Health (Berlin, Germany).²⁰ It offers artificial intelligence-powered symptom assessment technology to users. After users input their symptoms onto the platform, the tool uses AI to suggest a list of conditions that the user might have, and the probability associated with each suggested condition. In 2020, the SAA was shown to have market leading accuracy of condition-suggestion and urgency-advice accuracy at the same level as UK general medical practitioners.⁸ The SAA's reasoning engine infers disease probability estimations based on a representation of

medical knowledge, which is used to define a Bayesian network, on which approximate inference is carried out, and following which, information-theoretical methods are used to decide which questions to ask to the user.⁶ The SAA's knowledge base was built and reviewed by a large team of medical doctors with clinical experience in a curated process of knowledge integration from medical literature. It consists of disease models of all common conditions and several hundred rare diseases, including their corresponding symptoms and clinical findings. The disease models and their related symptoms are added to the knowledge base and modelled according to fundamental knowledge gained by their medical education and clinical experience, as well as evidence from medical textbooks and peer-reviewed medical literature.

The SAA's medical knowledge is expanded continuously following this standardised process and has been specifically optimised for a sub-Saharan Africa setting through a Foundation Botnar research grant (grant number 6270) to improve SAA applicability for LMIC populations. This has included symptoms/clinical-finding refinement based on additional attributes, for example, intensity or temporality and epidemiological data. These have been used to specify the prior probabilities of diseases to allow for correct probability estimations for this setting. It has also included optimisation of maternal and neonatal conditions, infectious diseases, non-communicable diseases, sexually transmitted infections, trauma-related injuries, mental health issues and neglected tropical diseases. In October 2019, the Ada SAA was made available at no cost to users in Swahili, via the Apple and Google app stores, and currently has over 92,000 users in Tanzania who have completed over 94,000 symptom assessments.

In a recent clinical investigation in a mental health use case of Ada in an outpatient clinic, the average completion time of an Ada assessment was 7.90 (SD 3.39) minutes and an average of 31.90 (SD 8.11) questions were asked.²¹ The Ada app is available both on Android and iOS devices in seven languages, and business-modified versions of the app can be found through various enterprise solutions with Ada business partners. A screenshot of the SAA used in the study in both English and Swahili is presented in [figure 1](#).

Study population and eligibility criteria

This study will assess children (2–119 months), adolescents (10–19 years) and other adults (20 years and above) who arrive at the study site. All patients who enter the clinic and are willing/able to provide consent will be included, with the exception of (1) Patients with severe injury/illness requiring immediate treatment, (2) Patients with traumatic injury (many of these patients require minimal anamnesis, and it is not rational to include them in a pilot study), (3) Patients incapable of completing a health assessment (e.g., due to illiteracy, mental impairment, inebriation, or other incapacity). Data from patients dropping out of the study or deviating from protocol will be excluded from analysis.



Figure 1 The left panel shows the Ada starting screen in English, and the right panel shows the screen in Swahili. After the starting screen, the user is guided through a series of questions about their presenting complaint and symptoms. The Ada app is currently available in seven languages.

Inclusion of patients will be monitored throughout the study in order to ensure recruitment of a study sample of patients with a comprehensive spectrum of symptoms, constellations and conditions: this is to ensure that this pilot study tests the performance of SAA on a broad range of scenarios, and that it not only provides detailed testing for the most commonly presenting patient scenarios. Study recruitment will be carried out to a target of enrolling between two and five patients for each of the following categories (including at least one adult and one child in each category): (1) Conditions related to abdominal pain or gastrointestinal issues, (2) Conditions related to the lower respiratory system, (3) Conditions related to the upper respiratory system, (4) Conditions related to mental health, (5) Conditions related to ophthalmology, (6) Conditions related to orthopaedics, (7) Conditions related to the cardiovascular system, (8) Conditions related to the genitourinary system, (9) Conditions related to ear, nose and throat (ENT), (10) Conditions related to the skin, (11) Conditions related to the female reproductive system and obstetrics, and (12) Conditions related to the neurological system. Once a total of five patients have been enrolled for a given category, no further patients will be included. There will be cases in which the presenting complaint does not match the condition category which the patient is ultimately diagnosed with; for this reason, the physician diagnoses will be aggregated on a dashboard recruitment adapted to optimise recruitment according to the categories listed above. The monitoring of this detailed recruitment is possible by the work of two study trackers, employed at the study site, who have nurse-midwife-level and clinical-officer-level medical training, respectively, and who can ensure the category tracking is followed; the recruitment is expected to become slower towards the end of the study because it is more difficult to enrol condition categories that still need to be filled.

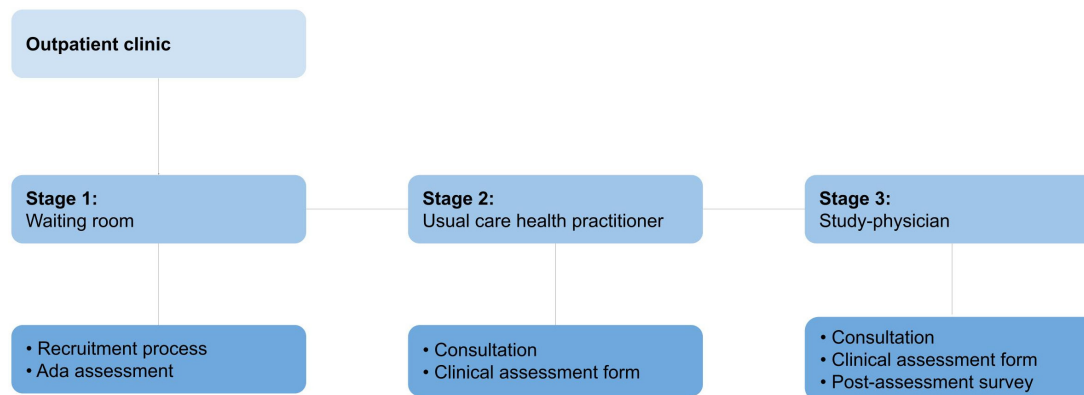


Figure 2 The patient journey in the study.

Description of study visits and assessment schedule

The pilot study will be preceded by a feasibility and process optimisation phase, in which 15 patients will be recruited. This phase is for the optimisation of general study procedures, of patient tracking and information recording in the busy clinic environment, and to determine if staff training has been adequate. If deficiencies in study process or staff training are identified, then these will be rectified, and a period of up to 2 weeks has been allowed in study planning for this. There will be no alteration in usual care for these patients and their data will not be analysed for the investigation of the study hypotheses. Following the feasibility phase, at least 50 patients will be recruited for the pilot study.

The patient's journey in the study will consist of three stages (see overview in [figure 2](#)).

Patient presenting to the clinic and Ada use (Stage 1)

Study recruitment staff will coordinate closely with hospital staff to determine when there is a potentially eligible patient in the waiting room, a system that has been successfully used to recruit patients in previous studies at this site. The study team will approach potentially eligible patients, provide details of the study, and obtain written informed consent. Parents/caretakers will be asked to consent for their children's participation. In addition, all children aged between 9 years and 18 years will be asked to provide an assent to their participation. The consent form will be in Swahili language (see online supplemental file 2 for the English version).

Each patient will be assigned a single study ID (this study ID is exclusive to the study and is not part of the usual care electronic health record). The SAA will be described to the patient, who will then use it independently to assess their symptoms. If the patient asks for assistance in SAA use, study staff will assist them and will record, on a modified Likert Scale, the degree of assistance provided. The results of the symptom assessment will not be shared with the patient or any of the health workers in the clinical setting.

Patient examination by usual care physician (Stage 2)

The patient will then proceed to usual care, which will either be a consultation with a clinical officer, an assistant

medical officer or with a medical doctor (here referred to collectively as 'usual care health practitioner'). A study-structured consultation form will be completed by the usual care health practitioner, which will either be a paper-based case report form (CRF) or tablet-based eCRF (using a REDCap application). Additionally, the usual care physician will fill in the standard hospital forms, collect vital signs and plan for investigations, as required.

Patient examination by study-provided physician (Stage 3)

All patients will then proceed to a consultation with a study-provided physician, who will also complete the appropriate structured consultation form as a tablet-based eCRF. The study-provided physician can only, at the end of their consultation, refer to the notes collected by the usual care health practitioner for effective management whenever needed, so as to avoid bias. After the patient has completed the full study process, the patient will be asked to complete a purpose-designed survey about the SAA. This survey can be found in the supplemental files as online supplemental file 3 and asks the patient questions surrounding how they liked the Ada assessment and to what extent using the Ada assessment before their doctor consultations made them approach the visit differently than they normally would.

Interventions

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

End points

Primary end points

The condition-suggestion accuracy and comprehensiveness of the SAA on a pilot level, evaluated against the gold standard differential diagnoses determined by the review panel, reported in the context of the accuracy of the usual care health practitioner. This current study is a pilot assessment of feasibility for the planning of a later study, and the number of subjects required to determine true accuracy and comprehensiveness will be included in a later study. In this pilot study, we determine a preliminary measure of accuracy and comprehensiveness.

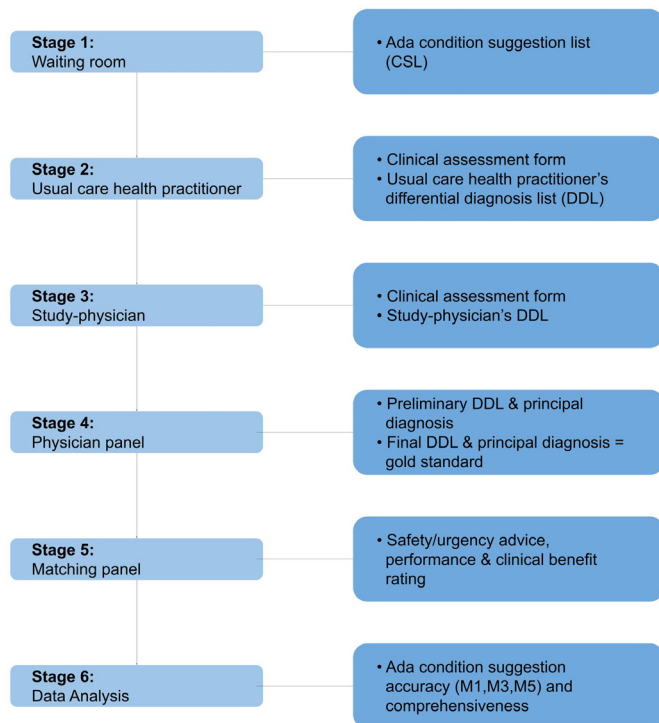


Figure 3 Data flow chart.

Additional data of interest

1. The urgency advice accuracy of the SAA, evaluated against the gold standard triage levels determined by the review panel, reported in the context of the accuracy of the usual care health practitioner.
2. Qualitative data on the usability, usefulness and acceptance of the SAA.

Measurement methods

The process of data collection and physician panel assessment consists of five stages (see overview in [figure 3](#)). The first three stages in this process have been described above (ie, (1) The patient SAA use; (2) The consultation with the usual care health practitioner; and (3) The consultation with the study-provided physician). The subsequent steps are:

4. Physician panel-generated differential diagnoses: A review process will be carried out by a physician panel in order to arrive at a gold standard differential diagnosis list and urgency-advice level for each case. The urgency-advice level and diagnoses from the usual care health practitioner and the study-provided physician and the pseudo-anonymised symptoms and medical history for each patient will be reviewed by two independent ‘reviewer’ physicians. Based on history and symptoms alone, they will assign a preliminary differential diagnosis list, a principal diagnosis and an overall urgency-advice level to each case, and, based on the full patient clinical file (including vital signs and results from medical examination and diagnostic tests), they will assign a final differential diagnosis list, principal diagnosis, and final view of the most appropriate urgency advice level. All diagnoses will be recorded



Figure 4 Gold standard urgency advice levels.

using the tenth revision of the International Statistical Classification of Diseases and Related Health Problems and urgency-advice level will be recorded on the scale shown in [figure 4](#).

5. Physician panel-generated matching of conditions: The reviewer physicians will judge the matching of usual care/attending study-provided physician diagnoses with their own diagnoses and carry out the same procedure for the SAA condition list. If there is disagreement between the two reviewer physicians on whether any condition is a match, then a third reviewer physician will make a definite decision, as in Gilbert, et al. and Semigran, et al. [^{8 18}]. The differential diagnosis list at each point in the patient’s clinical journey is a relevant point of comparison to the Ada assessment, but the differential diagnoses obtained before the addition of vital signs measurement, physical examination or additional diagnostic tests is the most important comparison point.

Data analysis will then be carried out on these lists (see section ‘Data analysis’ below). Patient questionnaires and usual care health practitioner questionnaires will be analysed and described using methods appropriate to modified Likert scale questionnaire data.²²

Risk-benefit assessment

As this is an observational (ie, non-interventional) study that does not pose any risk to the patient, there is no need for additional safety management, for example, a data monitoring committee. Patients requiring immediate medical care and clinically unstable patients are excluded from recruitment. There will be no delay in the diagnosis and treatment of any patients, since if they are called into their appointment before completing the SAA assessment, they will then be excluded from the

study process and analysis and instead proceed to their usual care. Study-enrolled patients will receive one extra consultation with the study-provided physician, which is highly unlikely to delay the patient's diagnosis or treatment, as this will generally require less than 10 minutes.

Data management and data safety

Data entry will take place according to the guidelines prepared in the study data management plan. For the paper CRFs completed by usual care healthcare practitioners, double data entry will be carried out (data will be entered by two operators separately). All consent forms will be paper based. All other data will be collected electronically. Data will be collected at the study sites through a secured local area network, which will allow data sharing on-site. A clinical trial Electronic Data Capture (EDC) system (REDCap), will be used for data capture. Study personnel will be trained on the system and be provided a unique username and password. Paper records will be kept in a locked cabinet in the facility and will only be accessed by study-specific personnel. At the point of EDC, the research assistant will verify the data and then commit it to the EDC. The data will then be automatically locked, and the research assistant will no longer have access to the data. Data collected from the study will be stored for a minimum of 3 years from the date of the last patient out by Muhimbili University of Health and Allied Sciences. Source-data verification will be conducted for 10% of the digitised usual care consultation notes.

Data analysis

Top-1, top-3 and top-5 accuracy (known as M1, M3 and M5) and comprehensiveness as defined by Gilbert *et al*⁸ of the SAA will be evaluated against the gold standard differential diagnoses, as will the accuracy and safety of the urgency-advice levels from the panel. Using the standard of M1, M3 and M5 in studies such as this one has been determined through its use in many other similar studies, as it shows the percentage of cases where the top-1, top-3 or top-5 condition-suggestion matches the gold standard main diagnosis.^{8 17 18 21} Data analysis will be conducted as described by Gilbert *et al*.⁸ Briefly, condition-suggestion accuracy and urgency-advice accuracy will be compared using descriptive statistics and tests appropriate for categorical data. χ^2 tests will be used to test whether the proportion of correct condition suggestions from the SAA, from the usual care medical practitioners and from the study-provided physicians are drawn from the same distributions. In case of a significant difference, two-sided post hoc pairwise Fisher's exact tests will be used to compare the SAA and practitioners. The usual care health practitioner rating will be strictly anonymised, as the purpose of the study is to compare the SAA to usual care and will not be used to audit individual usual care health practitioners. Condition categories as defined in the section 'Study population and eligibility criteria' will be broad and are used to gain a general overview on how the Ada app and doctors perform over different

categories. Analysing this will provide an important insight into the SAAs' strengths and limitations. There is limited research in clinical settings exploring how SAAs perform in different disease types of affected body systems.

Sample size and study timeline

The study was designed as a guide to a later larger trial, and in line with literature on pilot study design.^{23 24} Therefore, the sample size was estimated on the basis of having sufficient patients to assess accuracy and comprehensiveness on a pilot scale, survey completion rate, and to determine if there were any safety-related considerations that might be needed in a later larger study. Aspects that will be piloted are: (1) Trialling of new procedures and enabling power calculations intended to be used in a later single or multicentre randomised controlled trial; (2) Determining a pilot-based overview of accuracy and comprehensiveness for a comprehensive range of symptoms and conditions in varied age groups; (3) Establishing how many patients and/or healthcare professionals can be recruited and the feasible level of completed patient and physician questionnaires; and (4) Evaluating the general technical and logistic feasibility of a full-scale study, including issues of data collection and questionnaire design. This study is anticipated to start in July 2021 and the duration of patient recruitment to last for 2 months.

ETHICS AND DISSEMINATION

Ethical approval was received by the relevant ethics committee (EC) of Muhimbili University of Health and Allied Sciences with an approval number MUHAS-REC-09-2019-044 and the National Institute for Medical Research, NIMR/HQ/R.8c/Vol. 1/922. All amendments to the protocol are reported and adapted on the basis of the requirements of the EC. The results from this study will be submitted to peer-reviewed journals, local and international stakeholders, and will be communicated in editorials/articles by Ada Health.

Patient and public involvement

Patients were not directly involved in the development of the research question or study design; however, feedback from patient groups in related studies carried out at the study site have been used to help design this study and its interaction with patients.

DISCUSSION

Studies evaluating SAA accuracy

There is substantial literature on the performance of AI-powered SAAs in clinical contexts in high-resource settings, but none focus on LMICs. Many studies focus on vignettes rather than real patients for the evaluation of condition-suggestion and urgency-advice accuracy of SAAs. The condition-suggestion accuracy of the Ada SAAs underlying medical intelligence was evaluated in a retrospective study in medical records of patients diagnosed with rare diseases in Germany, in which information from

patients' health records was entered into an Ada prototype decision support tool. The tool was able to suggest the correct condition earlier than doctors in 56.3% of cases.²⁵ In a vignettes-based evaluation of SAAs in ENT-related illnesses, the Ada SAA was found to be the second best performing SAA.²⁶ There was also an evaluation of eight SAAs, including Ada, where UK General Practitioners (GP's) and the SAA's condition-suggestions and urgency-advice levels were compared against a gold standard in 200 clinically created vignettes spanning numerous condition types. It was found that condition-suggestion coverage was highly variable in SAAs, and that some SAAs did not offer a suggestion for many vignettes, but Ada had the highest condition coverage (99.0%). The top-3 suggestion accuracy for a GP was 82.1%±5.2%, whereas that for Ada was 70.5%; Ada was the best performing SAA of the eight SAAs evaluated. For safe urgency advice, tested GPs had an average of 97.0%±2.5%. For the vignettes with advice provided, only three SAAs, including the Ada SAA, had safety performance within 1 SD of the GPs.

Health tools and AI in Africa and other LMICs

It is widely recognised that there is high pressure on healthcare provision in LMICs, which in many regions has been heightened during the COVID-19 pandemic.¹⁵ Although there are no published studies on SAA use in sub-Saharan Africa, there are published evaluations of digital health applications, including AI-based tools. Individual AI-based tools have been used in sub-Saharan Africa to detect many health issues, including congenital heart defects, tuberculosis, blindness due to diabetes and many others.^{11 27 28} AI-based tools have not only been leveraged in the diagnostic process, but also to stratify risk, such as malaria risk and mortality in children due to acute infection.^{29 30} AI-based tools have also been used in sub-Saharan Africa to rapidly identify critically ill children and to facilitate timely intravenous antibiotic administration.³¹ These examples relate to AI-based tools with relatively narrow diseases or symptom areas of application; therefore, the current study differs, as the SAA evaluated is applicable across all common conditions and a wide range of rarer and tropical diseases.

Strengths and limitations of the current study

The study should provide a better understanding of local LMIC requirements for SAAs. In a systematic review, Chambers *et al* identified five limitations of published studies on the safety and accuracy of SAAs,¹ with most studies having several of these limitations: (1) Not being based on real patient data; (2) Not describing differences in outcomes between symptom assessment apps and health professionals; (3) Covering only a limited range of conditions; (4) Covering only uncomplicated vignettes; and (5) Sampling a young healthy population not representative of the general population of users of the urgent care system. The current study has been set up in such a way in order to reduce these limitations as much as possible: it will be based on real patient data, it will show the difference in outcomes between SAAs

and health professionals, and it will cover a wide range of conditions and age groups. It will also include patients with a range of health statuses, excepting those requiring urgent care or with chronic conditions. As this study will take place in an LMIC clinic waiting room, there will be many advanced cases presented alongside more straightforward cases. It should be noted that this is a pilot study and therefore is limited in the total number of recruited subjects. Through ensuring targeted recruitment of patients in specified disease categories, we ensure that the SAA is tested with a wide range of presenting scenarios, symptoms and diseases, and it avoids over-representation of common diseases and symptoms presentations in the study population. This is a strength of the pilot design and following this approach will also enable detailed exploration of a wide range of conditions and symptom presentations in different patient age groups in a later larger study.

Another strength of this study is that its general approach is based on other peer reviewed published studies. The SAA Mediktor was tested in a Spanish emergency department, with similar inclusion criteria: it included patients with medical problems which did not require emergency care. However, the study differs from this one, in that the SAA Mediktor study only included patients over 18 years.¹⁷ Unlike in the current study, the Mediktor SAA study app used the diagnosis of a single doctor as the gold standard diagnosis list. In order to achieve a greater degree of objectivity in the gold standard, we will use the three-physician-tie-breaker approach of Gilbert *et al* and Semigran *et al*.^{8 18} to determine the gold standard diagnosis list. Additionally, during the analysis phase, the studies on the Mediktor app excluded patients whose confirmed diagnoses were not modelled within Mediktor SAA's knowledge base, giving a biased and limited portrayal of the tool's accuracy. To achieve a more rigorous measure of accuracy, we will include all patients in analysis, regardless of whether their conditions are modelled in the Ada SAA medical knowledge base.

While some SAAs are limited by being based on rigid decision tree algorithms, this does not apply to the SAA evaluated in this study, as Ada's questioning is based on a dynamic approach, with adaptation to each point of new information. Ada uses Bayesian reasoning to ask the user questions after collecting demographic information, medical history and symptoms, in order to suggest possible conditions and urgency advice through interacting with the 'Ada Medical Knowledge'. Ada's questioning approach is similar to that of human doctors: the answer to the question before dynamically determines the next one asked, resulting in a manageable total number of questions asked and questions relevant to the patient's actual health state.

FUTURE PERSPECTIVES

As mentioned in this paper, this pilot study is part of a larger study which will more thoroughly look into all the questions still unanswered by this pilot study. The larger study will also be optimised to address any gaps that are found in this pilot study. Additionally, this study is part of an iterative process

of observational clinical testing and product development of the Ada SAA. This process is a feedback loop including pilot and larger clinical investigations, with insights being fed into product refinement in order to optimise Ada's safety and performance and to enable it to provide the best advice to individuals on their health.

Author affiliations

¹Ada Health GmbH, Berlin, Germany

²Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania

³EKFZ for Digital Health, Technische Universität Dresden, Dresden, Germany

Contributors EM, NS, HA, MMB, LO, MS, PB, ET, RV & SG contributed to the planning (study conception, protocol development). All the authors contributed to commenting on drafts of the protocol. SG is the guarantor for this work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding This study was funded by Foundation Botnar (grant number 6270).

Competing interests EM, HA, LO, MS, PB, ET and SG are employees or company directors of Ada Health GmbH and some of the listed hold stock options in the company. RV is a former employee of Ada Health GmbH. HA is a director of the Ada Health Foundation gGmbH. The Ada Health GmbH research team has received research grant funding from Foundation Botnar and the Bill & Melinda Gates Foundation.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Marcel Schmude <http://orcid.org/0000-0002-7519-1724>

Stephen Henry Gilbert <http://orcid.org/0000-0002-1997-1689>

REFERENCES

- Chambers D, Cantrell AJ, Johnson M, *et al*. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open* 2019;9:e027743.
- Mcdaid D, Park A-L. Online health: Untangling the web 2011.
- Van Riel N, Auwerx K, Debbaut P, *et al*. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open* 2017;1:bjgpopen17X100833.
- Aboueid S, Liu RH, Desta BN, *et al*. The use of artificially intelligent Self-Diagnosing digital platforms by the general public: Scoping review. *JMIR Med Inform* 2019;7:e13445.
- Meyer AND, Giardina TD, Spitzmueller C, *et al*. Patient perspectives on the usefulness of an artificial Intelligence-Assisted symptom Checker: cross-sectional survey study. *J Med Internet Res* 2020;22:e14679.
- Miller S, Gilbert S, Virani V, *et al*. 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:e19713.
- Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. *Lancet* 2018;392:2263-4.
- Gilbert S, Mehl A, Baluch A, *et al*. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? a clinical vignettes comparison to GPs. *BMJ Open* 2020;10:e040269.
- Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. *Med J Aust* 2020;212:514-9.
- Morse KE, Ostberg NP, Jones VG, *et al*. Use characteristics and triage acuity of a digital symptom Checker in a large integrated health system: population-based descriptive study. *J Med Internet Res* 2020;22:e20549.
- Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet* 2020;395:1579-86.
- World Health Organization. Tracking universal health coverage: 2017 global monitoring report, 2017. Available: <https://apps.who.int/iris/handle/10665/259817> [Accessed 28 Jun 2021].
- World Development Report. Governance and the law. world bank, 2017. Available: <https://www.worldbank.org/en/publication/wdr2017> [Accessed 28 Jun 2021].
- Weinberg J. Symptom tracker apps: the future of healthcare? *Raconteur* 2020 <https://www.raconteur.net/healthcare/symptom-tracker-apps/>
- Fry MW, Saidi S, Musa A, *et al*. "Even though I am alone, I feel that we are many" - An appreciative inquiry study of asynchronous, provider-to-provider teleconsultations in Turkana, Kenya. *PLoS One*;15:e0238806.
- Winn AN, Somai M, Fergestrom N, *et al*. Association of use of online symptom Checkers with patients' plans for seeking care. *JAMA Netw Open* 2019;2:e1918561.
- Moreno Barriga E, Pueyo Ferrer I, Sánchez Sánchez M, *et al*. [A new artificial intelligence tool for assessing symptoms in patients seeking emergency department care: the Mediktör application]. *Emergencias* 2017;29:391-6.
- Semigran HL, Linder JA, Gidengil C, *et al*. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
- Chan A-W, Tetzlaff JM, Altman DG, *et al*. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200-7.
- Health. Powered by Ada. Ada. Available: <https://ada.com/> [Accessed 28 Jun 2021].
- et alHennemann S, Kuhn S, Withhöft M. Can an App-Based symptom Checker accurately detect mental disorders: a comparative study in psychotherapy outpatients. *JMIR preprints*, 2021. Available: <https://preprints.jmir.org/preprint/32832> [Accessed 20 Dec 2021].
- Scheder-Bieschin J, Bluemke B, de BE. Improving emergency department patient-doctor conversation through an artificial intelligence symptom taking tool: an action-oriented design pilot study. *medRxiv* 2020.
- Moore CG, Carter RE, Nietert PJ, *et al*. Recommendations for planning pilot studies in clinical and translational research. *Clin Transl Sci* 2011;4:332-7.
- Thabane L, Ma J, Chu R, *et al*. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol* 2010;10:1.
- Ronick S, Hirsch MC, Türk E, *et al*. 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;14:69.
- Nateqi J, Lin S, Krobath H, *et al*. Vom symptom Zur diagnose - Tauglichkeit von Symptom-Checkern. *HNO* 2019;67:334-42.
- Bellefio V, Lim ZW, Lim G, *et al*. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health* 2019;1:e35-44.
- Thomford NE, Bope CD, Agamah FE, *et al*. Implementing artificial intelligence and digital health in resource-limited settings? top 10 lessons we learned in congenital heart defects and cardiology. *OMICS* 2020;24:264-77.
- Kabaria CW, Molteni F, Mandike R, *et al*. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar ES Salaam. *Int J Health Geogr* 2016;15:26.
- Kwizera A, Kisoorn N, Musa N, *et al*. A machine Learning-Based triage tool for children with acute infection in a low resource setting. *Pediatr Crit Care Med* 2019;20:1-30.
- Lee V, Dunsmuir D, Businge S, *et al*. Evaluation of a digital triage platform in Uganda: a quality improvement initiative to reduce the time to antibiotic administration. *PLoS One* 2020;15:e0240092.