Triage Accuracy and the Safety of User-initiated Symptom Assessment with an Electronic Symptom Checker in a Real-life Setting: Comparison Study

Ville Liu, Tuomas H. Koskela, Minna Kaila

Submitted to: JMIR Human Factors
on: December 02, 2023

Table of Contents

Triage Accuracy and the Safety of User-initiated Symptom Assessment with an Electronic Symptom Checker in a Real-life Setting: Comparison Study

Ville Liu[1] DMD, PhD Program; Tuomas H. Koskela[2] MD, PhD; Minna Kaila[3] MD, PhD

[1]Faculty of Medicine University of Helsinki Helsinki FI
[2]Department of General Practice Faculty of Medicine and Health Technology Tampere University Tampere FI
[3]Public Health Medicine Faculty of Medicine University of Helsinki Helsinki FI

Corresponding Author:
Ville Liu DMD, PhD Program
Faculty of Medicine
University of Helsinki
Ruusulankatu 21 B 32
Helsinki
FI

Abstract

Background: Previous studies have evaluated the accuracy of the diagnostics of electronic symptom checkers (ESCs) and triage using clinical case vignettes. National Omaolo Digital services (Omaolo) in Finland consist of an ESC for various symptoms. Omaolo is a medical device (based on Duodecim Clinical Decision Support EBMEDS) with a CE marking (risk class: IIa).

Objective: This study investigates how well triage performed by the ESC matches the triage of nurses among the chief symptom assessments available in Omaolo (anal region symptoms, cough, diarrhea, discharge from the eye or watery or reddish eye, headache, heartburn, knee symptom or injury, lower back pain or injury, oral health, painful or blocked ear, respiratory tract infection, sexually transmitted disease, shoulder pain or stiffness or injury, sore throat or throat symptom, urinary tract infection). In addition, this study assesses the accuracy, specificity, sensitivity and safety of the Omaolo ESC.

Methods: This is a clinical validation study in a real-life setting. It is a multicenter study in primary health care (PHC) centers across Finland. The included units were of the walk-in model of primary care, where no prior phone call or online contact was required. Upon arriving at the PHC center, users (patients) filled out the ESC questions, resulting in a triage recommendation, and after that a nurse assessed their triage. Findings concerning 877 patients were analyzed. The triage recommendations by the ESC were matched with the triage given by the triage nurse.

Results: The mean of the exact match for all symptom assessments were in 471out of 877, 53.7% (CI95 49.2%, 55.9%). The mean of the exact match or overly conservative but suitable (ESC's assessment was one triage level higher than the nurse's triage) for all symptom assessments were in 584 out of 877, 66.6% (CI95 63.4%, 69.7%). Safe assessments by the ESC accounted for 856 of 877, 97.6% (CI95 95.6%, 98.0%) of all assessments made. Concerning acute cases in which the nurse evaluated that a user needed to be treated urgently the ESC's exactly matched accuracy were in 244 out of 344, 70.9%, (CI95 65.8%, 75.7%). Sensitivity for the Omaolo ESC was 62.6% and specificity of 69.2%. Critical assessments were identified and further analyzed. In these 21 assessments, no indication was found that patient safety had been compromised.

Conclusions: This study was the first to assess the accuracy, specificity, sensitivity and safety of the Omaolo ESC in the Finnish PHC context. The findings indicate that the Omaolo ESC is safe to use when compared to the triage assessment of a nurse. Omaolo ESC has the potential to direct the patient user to the right place at the right time in terms of triage.

(JMIR Preprints 02/12/2023:55099)
DOI: https://doi.org/10.2196/preprints.55099

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Ville D. Liu[1], BM; Minna Kaila[2], MD, PhD, professor; Tuomas H. Koskela[3,4], MD, PhD, professor
[1] Faculty of Medicine, University of Helsinki, Helsinki;
[2] Faculty of Medicine, University of Helsinki, Clinicum, Helsinki;
[3] Faculty of Medicine and Health Technology, Tampere University, Tampere;
[4] Center of General Practice, Tampere University Hospital, Tampere, Finland

Triage Accuracy and the Safety of User-initiated Symptom Assessment with an Electronic Symptom Checker in a Real-life Setting: Comparison Study

Abstract

Background: Previous studies have evaluated the accuracy of the diagnostics of electronic symptom checkers (ESCs) and triage using clinical case vignettes. National Omaolo Digital services (Omaolo) in Finland consist of an ESC for various symptoms. Omaolo is a medical device with a CE marking (risk class: IIa), based on Duodecim Clinical Decision Support EBMEDS.

Objectives: This study investigates how well triage performed by the ESC nurse triage within the chief symptom list available in Omaolo (anal region symptoms, cough, diarrhea, discharge from the eye or watery or reddish eye, headache, heartburn, knee symptom or injury, lower back pain or injury, oral health, painful or blocked ear, respiratory tract infection, sexually transmitted disease, shoulder pain or stiffness or injury, sore throat or throat symptom, urinary tract infection). In addition, the accuracy, specificity, sensitivity and safety of the Omaolo ESC were assessed.

Methods: This is a clinical validation study in a real-life setting, performed at multiple primary health care (PHC) centers across Finland. The included units were of the walk-in model of primary care, where no prior phone call or online contact was required. Upon arriving at the PHC center, users (patients) answered the ESC questions, and received a triage recommendation, where after a nurse assessed their triage. Findings on 877 patients were analyzed by matching the ESC recommendations with triage by the triage nurse.

Results: Safe assessments by the ESC accounted for 856 of 877, 97.6% (CI95 95.6%, 98.0%) of all assessments made. The mean of the exact match for all symptom assessments were in 471out of 877, 53.7% (CI95 49.2%, 55.9%). The mean of the exact match or overly conservative but suitable (ESC's assessment was one triage level higher than the nurse's triage) for all symptom assessments were in 584 out of 877, 66.6% (CI95 63.4%, 69.7%). When the nurse concluded that urgent treatment was needed, the ESC's exactly matched accuracy were in 244 out of 344, 70.9%, (CI95 65.8%, 75.7%). Sensitivity for the Omaolo ESC was 62.6% and specificity of 69.2%. Twenty one critical assessments were identified for further analysis: there was no indication of compromised patient safety.

Conclusions: The primary objectives of this study were to evaluate the safety and to explore the accuracy, specificity, and sensitivity of the Omaolo ESC. The results indicate that the ESC is safe in a real-life setting when appraised with assessments conducted by triage nurses. Furthermore, the Omaolo ESC exhibits potential to guide patients to appropriate triage destinations effectively, helping them to receive timely and suitable care.

Keywords:

1. **Triage**
Unique ID
D014218
2. **Symptom Assessment**
Unique ID
D063189
3. **Self Care**
Unique ID
D012648
4. **Health Services Accessibility**
Unique ID
D006297
5. **Telemedicine / eHealth**
Unique ID
D017216
6. **Health Services Research**
Unique ID
D006302
7. **Public Health**
Unique ID
D011634
8. **Validation study**
Unique ID
D023361
9. **Primary Health Care**
Unique ID
D011320
10. **Remote consultation**
Unique ID
D019114

Introduction

Background

Seeking information online regarding medical symptoms is a common and well-known phenomenon among people and patients worldwide [1-4]. Usually, the general public searches online for symptoms associated with their medical condition before receiving a medical

diagnosis. This includes websites of support groups, patients' written blogs, websites created by editors of popular media, governmental sites, and AI interfaces. However, self-diagnostic web-based sources may be of varying quality in terms of reliability, with misleading information and possibly false advertising [5-6].

To address these problems, digital healthcare applications are spreading online, including self-diagnosis tools and electronic symptom checkers (ESC) [7-10]. These are meant to provide solutions and information to the user seeking to learn more about symptoms or a condition they have or think they might have. In cases where access to human healthcare experts may be limited, telehealth services have tremendous promise for transforming the provision of healthcare services [11]. Conversely, studies find that healthier users use digital services more often than others and are also more likely to be younger, female, and more highly educated, and to have higher income levels [10,12-13].

Based on the user's input, ESCs use algorithms to make diagnostic suggestions, offer advice on what action to take, and help in identifying the relevant condition. This is medical triage, and it involves directing patients to the most suitable location within an appropriate time frame. In clinical practice, triage assessment and guidance are usually done by health care professionals either over the phone or face to face, e.g., at a healthcare center [14]. Triage takes up a lot of professionals' time and its quality varies. Therefore, even the partial digitalization of triage in healthcare organizations could increase service uniformity, enhance efficiency, and free up working hours [15-16]. This inherently requires that health organizations and teams reorganize their workflows and work distributions to support clinical processes [17-18].

The Omaolo ESC questionnaires and algorithms in use are based on research evidence, probabilities, and expert opinions as to whether the condition described is mild and self-limiting. In terms of urgency, an assessment is made on how soon the condition would worsen without treatment or whether it requires the intervention of a health care professional. However, as with clinical decision making in general, making an accurate diagnosis requires user-provided information, clinical examinations, various diagnostic tests, and potential consultations with other health care professionals [19-20].

Previous studies have evaluated the accuracy of the diagnostics of ESCs and triage using clinical case vignettes [21-27]. Variation exists between different ESCs and the conditions being assessed, including the triage capabilities [8, 27-29]. In some studies, the diagnostic accuracy of clinicians has been shown to be superior in both primary and specialized health care when compared to ESC tools [22-23]. These studies have shown that users may be referred to self-care even if they need professional help, and users for whom self-care would suffice are referred to unnecessary counseling. There are risks and the potential for error in ESC-based triage [8, 21-29]. In particular, self-care guidance should be limited to cases where it is safe and appropriate. There is currently a limited amount of evidence available on the impact of ESCs on seeking treatment with real-life users [12, 30]. However, respondents have been satisfied with the ESC services they use [13, 31-32].

A study comparing the accuracy of physicians' and computer diagnostics found that physicians listed the correct diagnosis first more often across all study vignettes compared with ESCs (79.1-65.3% vs. 40.5-24.3% $P < .001$) as well as in the top 3 diagnoses listed (84.3% vs 51.2%, $P < .001$) [24]. There is limited evidence of live clinical patient safety hazards associated with the use of ESCs, as safety has mainly been evaluated with the use of clinical vignettes [7-8, 21, 23-24,26]. When comparing artificial intelligence and human doctors concerning triage and diagnosis, some AI systems were able to provide triage and diagnostic information on a level of clinical accuracy and safety comparable to human doctors [8, 23, 33]. However, ESCs on average make the user's triage more sensitive to the need for more urgent care than the user would need [22, 34].

The seamless integration of ESCs into the broader healthcare triage process is crucial for achieving their intended goals, such as preventing emergency room overcrowding and providing more accurate symptom assessment and triage for citizens. ESCs can offer citizens a preliminary triage level for their symptoms before contacting healthcare services [35]. Additionally, ESCs and eHealth applications can serve an educational purpose by providing users with structured, research-based disease and treatment information that is easily accessible [36-37]. From a clinical perspective, the ability to accurately identify cases where self-care suffices is paramount in assessing an app's utility in preventing overcrowding and the "unnecessary use of healthcare services" [38].

Description of the Omaolo ESC

Omaolo is a national web-based service for health care and social welfare. The purpose of Omaolo is to promote the health and well-being of citizens. Omaolo supports self-care and helps people to contact public health care professionals, if necessary. Omaolo is a medical device with a CE marking, manufactured by government-owned DigiFinland and Duodecim Publishing Company Ltd. Omaolo was granted a CE certificate in accordance with the requirements of EU Regulation 2017/745 (MDR) in May 2022 [39]. The aim of the Omaolo ESC is to identify, based on an assessment of alarm symptoms and other pre-existing conditions, situations that require immediate or urgent assessment, and to conduct follow-up examinations and treatment without delay in situations where conservative treatment may lead to complications. The questionnaires and the algorithms the ESC uses are based on evidence and legal requirements [40].

The ESC operates as follows [41]: The user initially receives reliable information about the symptom (articles from the Health Library Duodecim) and a short summary. If unable to decide on the course of action needed, the user can answer the ESC's questions. The ESC will then suggest the estimated needed treatment and its urgency. The results of the completed survey made by identified users are saved and prompted to be sent to a regional healthcare professional through Omaolo. The ESC algorithm initially seeks to identify alarming symptoms and then prompt the user to contact the nearest emergency department immediately. The idea is to identify situations where a professional assessment is necessary and to determine the urgency of the assessment. The user is encouraged to consider whether they may have symptoms that are not covered by the information or survey provided. As the questionnaires might not cover all possible situations that could be due to other illnesses, treatments, or other causes that the user may have, the following help text is displayed to the user at the end of the query: "If you have symptoms that have not been covered in the survey or other illnesses or medications that you think affect your need for treatment, contact your PHC provider or, in an emergency, the nearest emergency department." [41]. Description of the Omaolo service is described in detail in the study protocol [40].

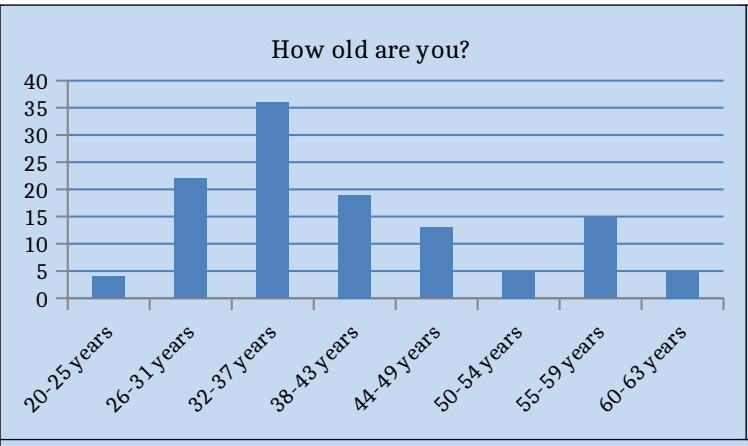
Objectives

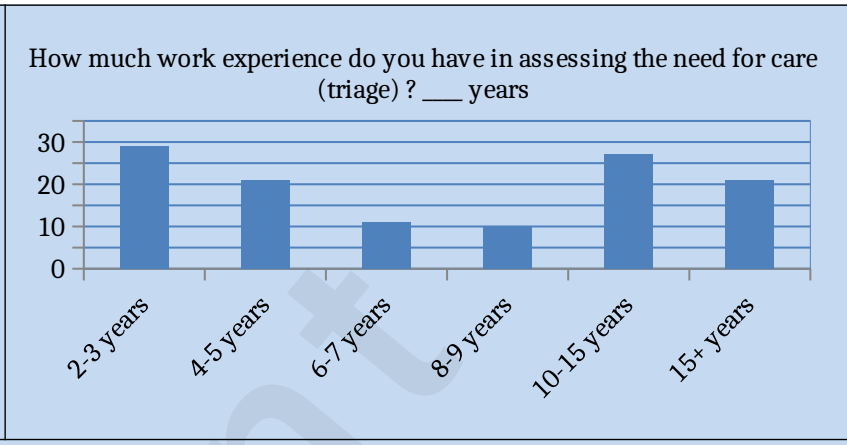
Our hypothesis is that the Omaolo ESC assessments are safe to use compared to the assessment of a triage nurse.

The aim is to study the clinical validity of the Omaolo ESC and to evaluate its exact triage accuracy, specificity, sensitivity and safety. These parameters can be used to determine if Omaolo ESC can direct the right patient user to the right place at the right time. The main research question was: How well does triage by the ESC match the triage of a nurse? [40], Appendix 1.

Methods

Study design and setting





How did the user arrive at the reception (walk-in or via telephone contact)?
Did you consult a doctor to assess the user's need for treatment (triage)?
What was the most significant thing (observation, symptom, or discovery) that influenced your decision making?
Do you feel the need to change your assessment of the need for treatment (triage) after seeing the responses and recommendations of the ESC? (Yes / No)
If yes, why did you change the assessment of the need for treatment (triage) after seeing the responses and recommendation of the ESC?
If you feel it is necessary to change your triage assessment, reselect where the user should be referred to according to the classification terms of the ESC recommendation.

The study setting was primary health care (PHC), and the data were collected at 18 PHC centers nationwide in Finland. The data were collected between June 1, 2018 and December 31, 2020. The study used the version of the Omaolo ESC that were in use in 2018. Each participating organization provided a study space where it is practically possible for the users (patients) to complete the ESC questionnaire and a nurse with at least two years' experience of triage nurse work in primary care to perform triage (Table 1).

Table 1. Distribution of triage nurse's age and work experience, and the nurse's form questions.

A total of 119 individual nurses took part in this study. A nurse's average age was 40 years (median 37 years, IQR = 14 years) and their average amount of work experience in triage 9.7 years (median 7 years, IQR 9.5 years).

Three in every ten recruited patient users arrived via walk-in at the PHC centers, and the rest first contacted their center via telephone. Upon physically arriving at the PHC centers the patients were asked if they were willing to participate in the study.

The patients answered the ESC when they arrived at the PHC center (on arrival at the center, they also filled out a consent form). Filling out the questions of the ESC was done in a separate quiet space without the research assistant interfering. Next, a triage nurse made a triage assessment of the same patient and filled out the study questionnaire related to the assessment of the patient's triage. The triage nurses did not get to know the result of the ESC until they had assessed the patient's condition themselves. After completing their questionnaire, the triage nurses finally got to see the result of the ESC triage concerning the same patient. Based on that result, the triage nurses filled out another questionnaire inquiring whether they felt it necessary to change their own assessment-based action recommendation after seeing the action recommendation of the ESC. The organization also ensured that the patient population

of the study remained unscreened. No user identifying age or gender data were collected for this study. [40] (Figure 1)

Assessment of ESC coverage and triage suggestion levels

The results of the ESC triage and the assessments of nurses were analyzed from the completed study forms. Each assessment was first analyzed individually and the results concerning particular symptom assessments were combined.
The MSAH (Ministry of Social Affairs and Health) has provided a practical classification for levels of emergency, and THL, the Finnish Institute for Health and Welfare referral class classification with codes [42]. (Figure 2, Figure 3 and Figure 4)
These classifications levels are as follows ranging from P0 to P4 and L2 to L4:

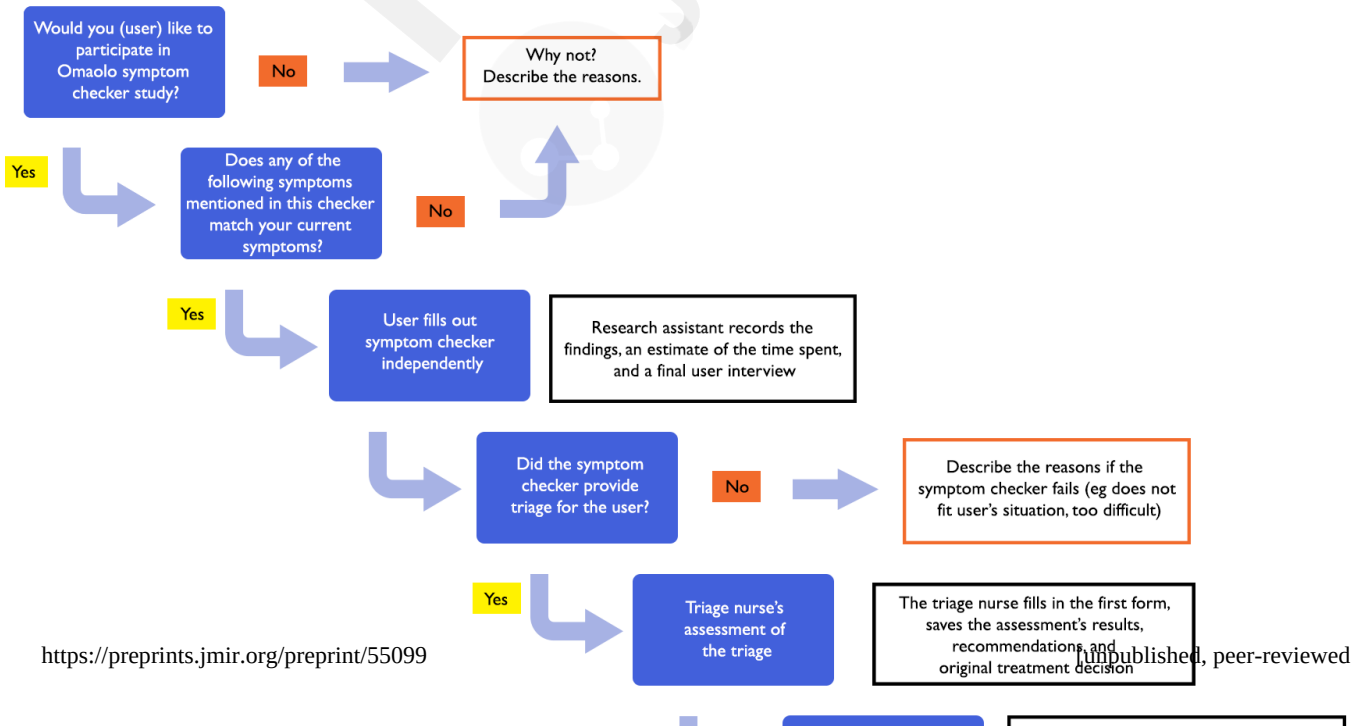
- P0. Treated immediately, contact emergency center [within 0h]
- P1. Contact emergency center services [within 2h]
- P2. Treated at an emergency center, but not during night [within 10h]
- P3. May be referred to a doctor during office hours [within 24h]
- P4. May be referred to a doctor during office hours [within 72h]
- L2-L4. Start self-care (with instructions)

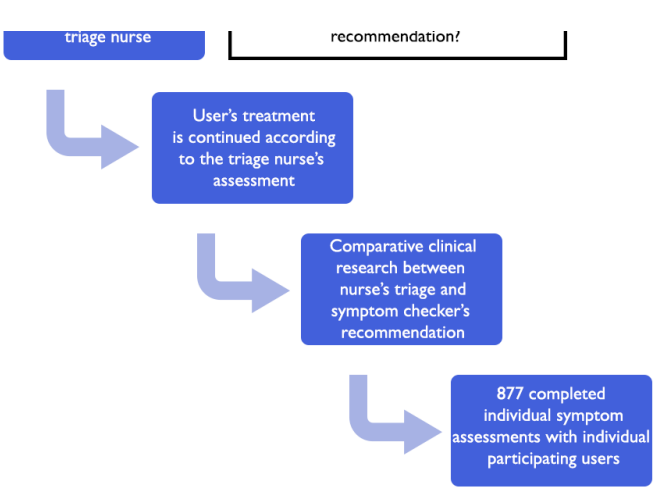
Making use of the MSAH emergency levels, the Omaolo ESC and nurse triage were matched as follows: exactly matched (ESC's and triage nurse's triage were the same), overly conservative but suitable (ESC's assessment was one triage level higher than the nurse's triage), safe but under conservative (ESC's triage level was one triage level lower than the nurse's triage), overly conservative (ESC's triage level was two levels higher than the nurse's triage), and potentially unsafe (triage nurse assessed as urgent or an on-call duty but ESC assessed as non-urgent or self-care).

Assigning accuracy gold standards to the triage nurse and the ESC

The nurse triage assessment is the gold standard, to which the recommendations of the ESC were compared. An assessment was also defined as potentially unsafe if a case was assessed by the nurse as urgent or an on-call duty but assessed by the ESC as non-urgent or self-care. These assessments were further investigated to ensure patient safety.

Figure 1. Flow chart showing the collection of the symptom checker and nurse triage data

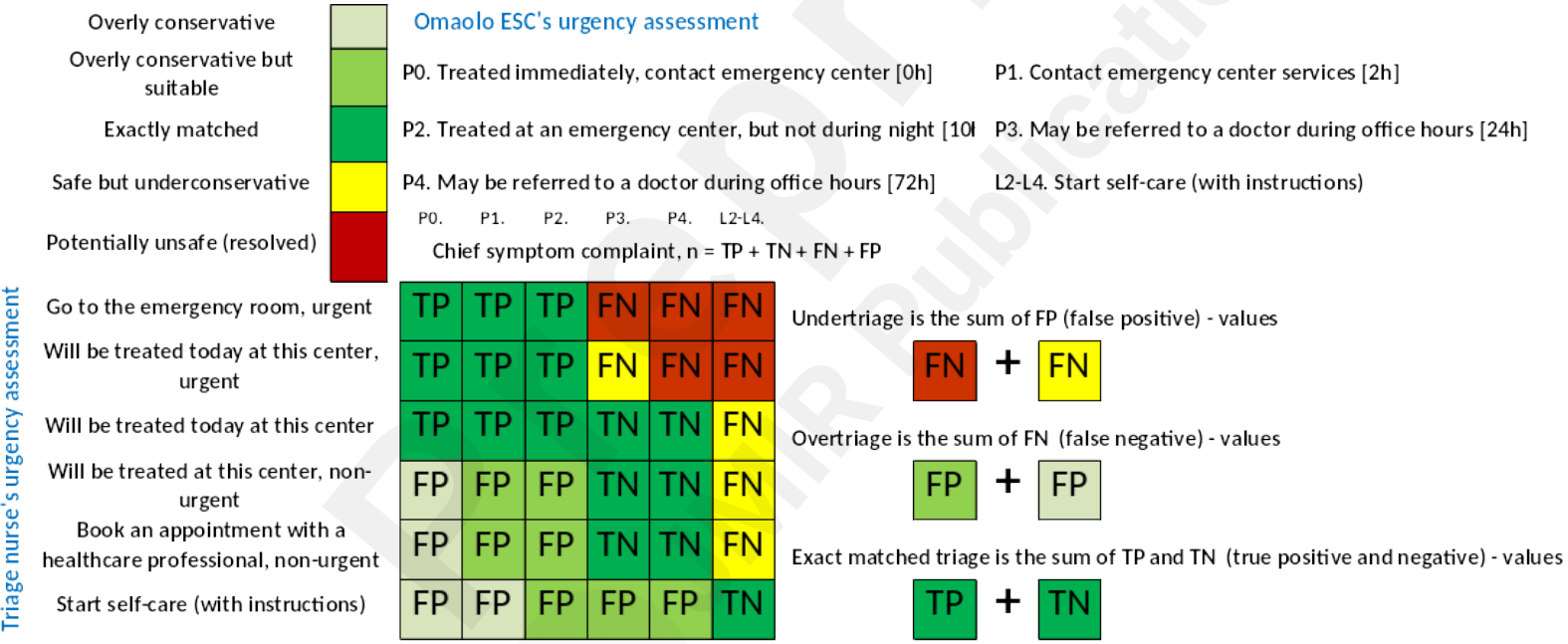




hypothesis is that the Omaolo ESC is safe to use [40].

We set the safe symptom assessment based on a previous study, as at most one level of urgency less urgent than the assessment of a nurse and the performance of the ESC as 97% safe accuracy across all completed assessments [23]. We also assumed that misdiagnosis by physicians occurs in approximately 5% of outpatients [43]. We estimated the required sample size by assuming the given range of safe advice at 97% and using a 95% confidence level, and we computed the confidence interval estimate for the true proportion of safe ESC assessments.

Figure 2. An example showing how the triage comparison chart (confusion matrices) in Figure 3 were constructed with color-coded differences in triage levels (overly conservative, overly conservative but suitable, exactly matched, safe but underconservative and potentially unsafe (resolved)). Matching rows (triage nurse) to their respective columns (symptom assessment) results in a safety assessment. P0-P4 = classification of emergency care criteria. L2-4 = referral urgency classes. The columns show how often, by ESC, the symptom assessment was overestimated, underestimated, or accurately matched compared to the assessment made by the triage nurse and the decision made on further referral.



Triage comparison charts (confusion matrices) are created in which true positive (TP) represents the outcome where the model correctly predicts the positive class (condition is detected when present). True negative (TN) is the outcome where the model correctly predicts the negative class (condition is not detected when absent). False positive (FP) represents the outcome where the model incorrectly predicts the positive class (condition is detected when absent). False negative (FN) is the outcome where the model incorrectly predicts the negative class (condition is not detected when present). These values are crucial components in calculating the positive predictive values (PPV), negative predictive values (NPV) and Matthews correlation coefficient (MCC). The PPV, NPV and MCC formulas are given as:

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

These values are typically extracted from a confusion matrix, as illustrated in Figure 3 and presented in Table 3 of the study.

The Matthews correlation coefficient ranges from -1 to +1, where ±1 indicates perfect agreement or disagreement, and 0 indicates no relationship.

Ethical considerations

This study was reviewed by the Pirkanmaa hospital district's ethics committee (ETL-Code:R18126), and a regional permission was in addition granted by each participating organization, all according to the regulations of the University of Tampere.

The most significant ethical issue related to the research setting is that the user's participation in the research does not affect their chances of receiving timely treatment. All users who fill out the symptom checker will be forwarded to an appointment with an experienced nurse. Denial of treatment for users who refuse to participate in the study was strictly prohibited.

When users (patients) were recruited for this study, the research assistant informed the user about the study, distributed the study information sheet, and then asked if the user was willing to participate in the study. If, after being informed, the user was willing to participate in the study, they were asked to sign a consent form in which the user acknowledged that they have received sufficient information about the study and agreed to participate in it. The user was given an information sheet about the study with contact information in case the user wanted to ask more about the study. The user was paid no amount of compensation.

The patient user's consent form was disconnected from the response form with a personal identification code, that is, the users were completely anonymized. No medical record data were collected or combined with research forms. The users could withdraw their consent to the study at any time. However, the completed forms cannot be destroyed after the data collection, because the consent form containing the personal data does not have the identification code to identify the appropriate study forms to be destroyed.

Results

A total of 877 patient user assessment cases were successfully collected. No patient user identifying age or gender data were collected for this study.

Quantifying and comparing the levels of urgency and triage accuracy

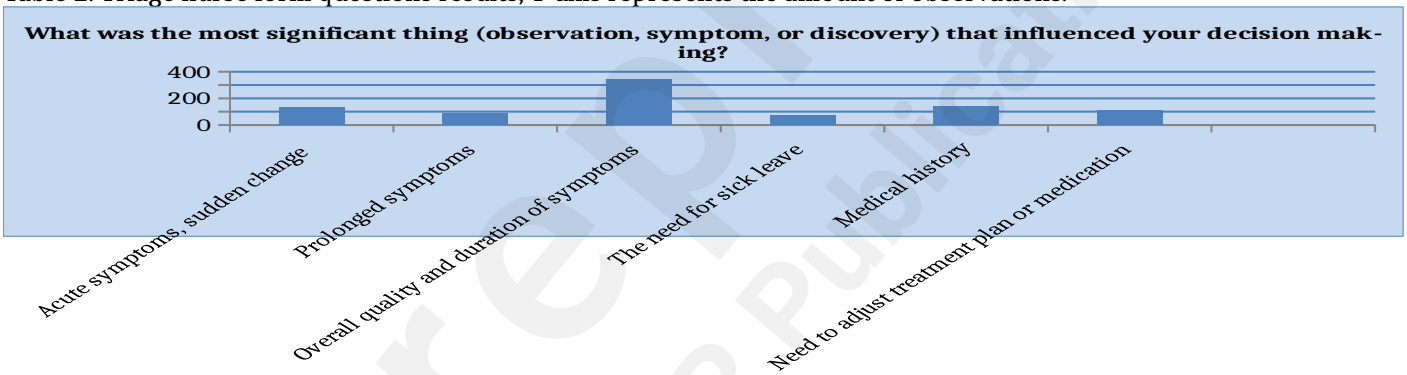
The ESC's and nurse's triage were exactly matched in 471out of 877, 53.7% (CI95 49.2%, 55.9%) of the cases in all symptom assessments. Considering ESC's individual main symptom triage suggestions the most exactly matched assessments were found for shoulder pain, stiffnes or injury (68%), dental or oral symptom or trauma (60%) and sexually transmitted disease (59%).

The mean exactly matched or overly conservative but suitable for all symptom assessments were in 584 out of 877, 66.6% (CI95 63.4%, 69.7%). Safe assessments of ESCs accounted for 856 of 877, 97.6% (CI95 95.6%, 98.0%) of all assessments made (Table 3, Figure 2, Figure 3 and Figure 4). Concerning acute cases in which the nurse evaluated that a user needed to be treated urgently the ESC's exactly matched accuracy were in 244 out of 344, 70.9% (CI95

65.8%, 75.7%) and in cases whether medical care should be sought or self-care is sufficient matches were found in 351 out of 533, 65.9% (CI95 61.7%, 70.0%). Sensitivity for the Omaolo ESC was 62.6% and specificity of 69.2%. The proportions of evaluations occurred at a ratio of 100 suitable to 25 overtriage to 22 undertriage. The overly conservative triage (overtriage) suggestions made by the ESC occurred most often for respiratory tract infection in 46 out of 104 (44%), heartburn in 8 out of 24 (33%), and headache in 15 out of 41 (37%). (Figure 4)

The question “Do you feel the need to change your assessment of the need for treatment (triage) after seeing the responses and recommendation from the ESC? (Yes / No)” was answered with the “Yes” option in 19 out of 877 assessments. In answering the follow-up question “If yes, why did you change the assessment of the need for treatment (triage) after seeing the responses and recommendation by the ESC?”, the nurses frequently stated that the symptoms they were told by the user did not match with the ones the user had stated while filling the ESC (Table 2 and Appendix 2). In the last section, in the case of an affirmative answer, the path of changing the triage ended with “If you feel it is necessary to change your triage assessment, reselect where the user should be referred to according to the classification terms of the ESC recommendation.” In these 19 assessments, the nurses were found to have chosen a triage suggestion closer to that of the ESC assessment (see Appendix 2 for details). In 80 cases across all completed 877 assessments, the triage nurse consulted a doctor in assessing the triage.

Table 2. Triage nurse form questions results, Y-axis represents the amount of observations.



Symptom	The number of completed matched assessments	Percentage of safely matched advice symptom assessments (%, 95% CI)	Exactly matched symptom assessments (%, 95% CI)	Positive predictive value (PPV), precision, %	Negative predictive value (NPV), %	Matthews Correlation Coefficient (MCC)
Anal region symptom	41	100 [91.4-100]	53.7 [37.4-69.3]	14.3	74.1	-0.133
Cough	71	95.8 [88.1-99.1]	50.7 [38.6-62.8]	52.2	48.1	0.004
Dental or oral symptom or trauma	62	85 [74.2-93.1]	59.7 [46.5-72.0]	62.9	57.7	0.204
Diarrhea	21	95.2 [76.2-99.9]	45.5 [24.4-67.8]	0	66.7	-0.37
Discharge from the eye, watery or reddish eye	65	100 [90.0-99.6]	58.5 [45.6-70.6]	66.7	28.6	-0.036
Headache	41	100 [91.4-100]	53.7 [37.4-69.3]	50.0	66.7	0.138
Heartburn	24	100 [85.8-100]	37.5 [18.8-59.4]	26.7	55.6	-0.233
Knee symptom or injury	55	100 [93.5-100]	56.4 [42.3-69.7]	36.8	66.7	0.035
Lower back pain or injury	65	100 [94.5-100]	56.9 [44.0-69.2]	46.7	80.0	0.253

Painful or blocked ear	81	100 [95.6-100]	58.0 [46.5-68.9]	20.0	49.0	-0.115
Respiratory tract infection	104	94.2 [88.0-97.9]	41.3 [31.8-51.4]	40.2	45.4	-0.118
Sexually transmitted disease	39	100 [91.0-100]	59.0 [42.1-74.4]	50.0	61.3	0.093
Shoulder pain, stiffness or injury	47	100 [92.5-100]	68.1 [52.9-80.9]	42.9	72.5	0.12
Sore throat or throat symptom	101	98 [93.0-99.8]	54.5 [42.3-62.5]	48.3	78.6	0.19
Urinary tract infection	60	100 [94.0-100]	51.7 [38.4-64.8]	57.1	46.9	0.04
In total	877	97.6 [96.4-98.5]	53.7 [50.3-57.1]	49.4	59.5	0.089

Absolute Magnitude of the Matthews Correlation Coefficient	**Interpretation**
±0.00-0.10	Negligible correlation
±0.10-0.39	Weak correlation
±0.40-0.69	Moderate correlation
±0.70-0.89	Strong correlation
±0.90-1.00	Very strong correlation

Table 3. Results of matching the nurse triage (gold standard) with the Omaolo ESC recommendation in 15 different symptoms. The Omaolo ESC recommendation is defined as safe if the critical condition was not met as assessed by the nurse as urgent or on-call duty but assessed by the ESC as non-urgent or self-care. Confidence interval range in the column reporting the number of completed assessments is assumed at safe advice of 97% and using 95% confidence level. The table also shows positive (PPV) and negative (NPV) predictive values following Matthews Correlation Coefficient (MCC).

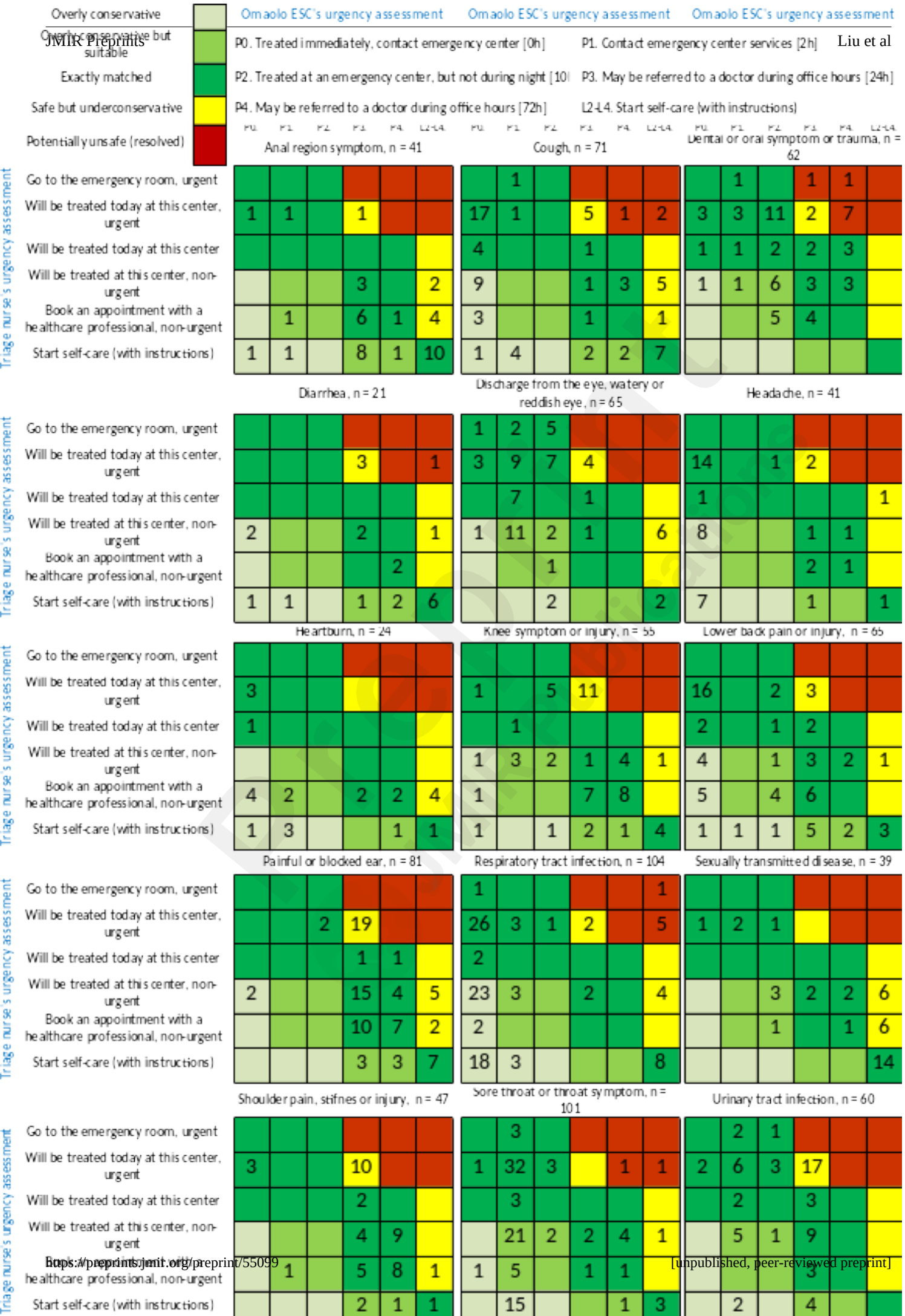

Overly conservative
Overly conservative but suitable
Exactly matched
Safe but underconservative
Potentially unsafe (resolved)
Omaolo ESC's urgency assessment
P0. Treated immediately, contact emergency center [0h]
P1. Contact emergency center services [2h]
P2. Treated at an emergency center, but not during night [10
P3. May be referred to a doctor during office hours [24h]
P4. May be referred to a doctor during office hours [72h]
L2-L4. Start self-care (with instructions)
Triage nurse's urgency assessment
Go to the emergency room, urgent
Will be treated today at this center, urgent
Will be treated today at this center
Will be treated at this center, non-urgent
Book an appointment with a healthcare professional, non-urgent
Start self-care (with instructions)
Anal region symptom, n = 41
Cough, n = 71
Dental or oral symptom or trauma, n = 62
Diarrhea, n = 21
Discharge from the eye, watery or reddish eye, n = 65
Headache, n = 41
Heartburn, n = 24
Knee symptom or injury, n = 55
Lower back pain or injury, n = 65
Painful or blocked ear, n = 81
Respiratory tract infection, n = 104
Sexually transmitted disease, n = 39
Shoulder pain, stifnes or injury, n = 47
Sore throat or throat symptom, n = 101
Urinary tract infection, n = 60

Figure 3. Triage comparison chart (confusion matrices) with color-coded differences in triage levels .

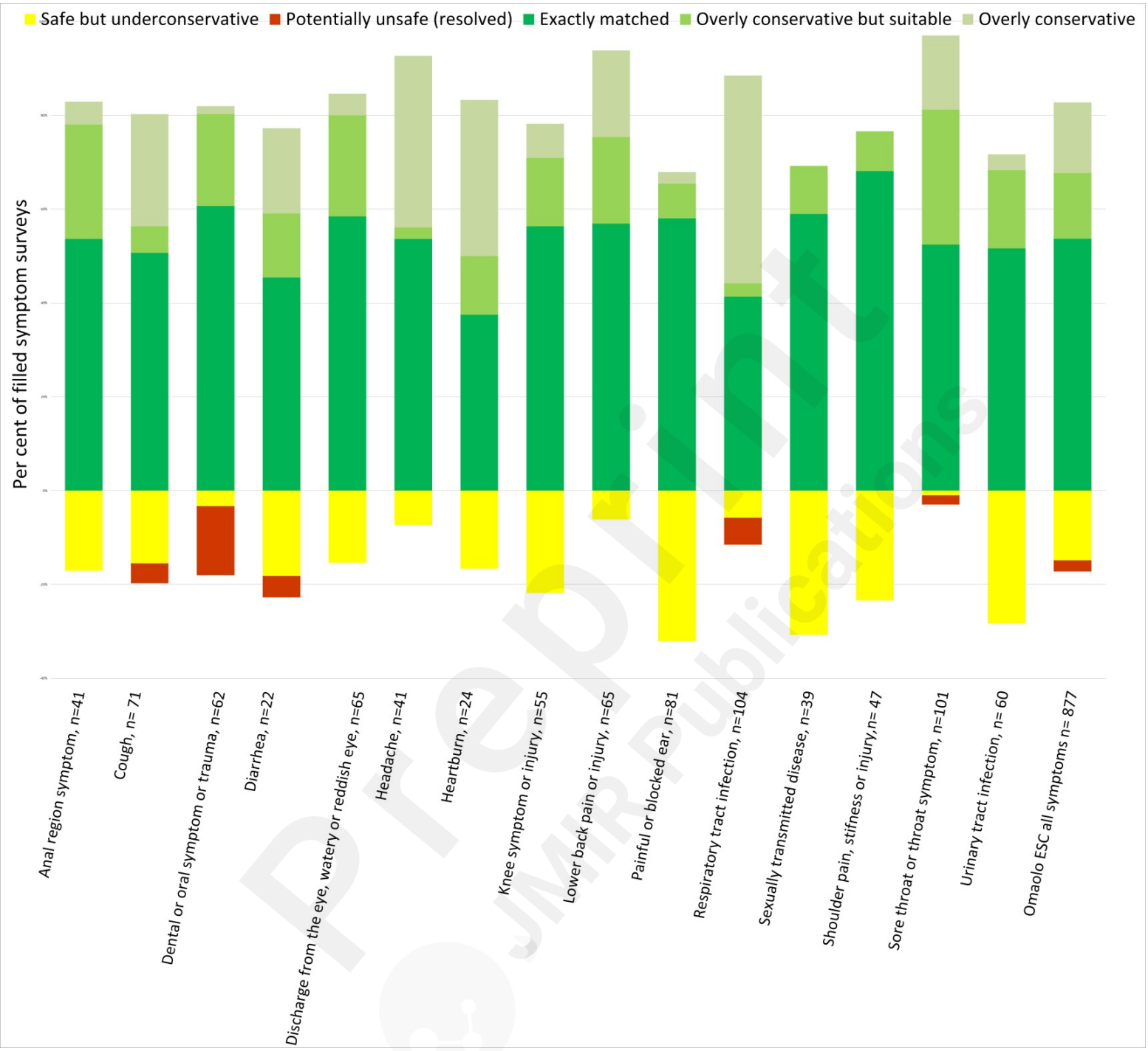


Figure 4. The individual bar heights (Y axis, %) reflect the proportional correspondence and accuracy levels of the ESC triage with the nurse's triage (overly conservative, overly conservative but suitable, exactly matched, safe but under conservative and potentially unsafe (resolved)). Individual ESCs are depicted on the X axis.

The analysis of critical assessments

Twenty-one critical assessments were identified for further analysis. Details are given in Appendix 3. Further analysis showed that there were no indications that patient safety was endangered. In Figure 3 and Figure 4 these patient user cases are marked as "Potentially unsafe (resolved).

Discussion

Principal Results and Findings

The findings suggest that while exact matches of the Omaolo ESC recommendations with the gold standard, nurse triage, occurred in just over half of the cases, nearly all cases were evaluated as safe, with urgency levels being at most one level less urgent compared to the nurses' triage. Concerning acute assessments, an exact match was found in nearly three out of four cases. This study assessed for the first time the safety of the Omaolo ESC within the Finnish PHC context.

Comparison to Prior Work

In a systematic review, the diagnostic accuracy and the triage-making abilities of ESC services such as Ada, Babylon, Buoy, K Health, Mediktor, Symptomate, and Your.MD were compared to those of general practitioners assessment using clinical case vignettes. The average safe operating recommendations ranged from 90.1% ±7.4% [23]. By contrast, the general practitioner's percentage of safe advice stood at 97.0% ±2.5% [23].Comparatively, the proportion of safe Omaolo ESC assessments across all investigated cases was 97.6% using similar methods for safety assessment. These findings underscore the safety of Omaolo ESC compared to assessments by experienced nurses, particularly notable given the real-life setting of our study. By comparing Omaolo ESC's accuracy to internationally reported results, we can gauge its overall performance and capabilities [8, 22-24, 44].

A study examining Ada's use by 378 "walk-in" patients in urgent care compared its triage accuracy with that of a triage nurse using the Manchester Triage System, conducted under similar circumstances as the Omaolo ESC's triage accuracy study. Ada exhibited an undertriage rate of 8.9% and an overtriage rate of 57.1%. Out of 378 participants, 344 (91%) were triaged identically or more conservatively, while 34 (8.9%) were undertriaged by the app. The assessment was deemed safe in 94.7% (358/378) of patients when compared to the Manchester Triage System assessment. In the Omaolo ESC study involving 877 users, 726 (83%) were triaged identically or more conservatively, with 151 (17%) being undertriaged. Notably, Omaolo ESC exhibited a 29% overtriage rate, with evaluations occurring at a ratio of 100 suitable evaluations to 25 overtriages to 22 undertriages. Compared to Ada, in the Omaolo study overtriage rates were lower.

A recent systematic review concluded that the median accuracy of studied apps in determining the necessity of emergency care was 80% (IQR 74.6%-86.8%) [20]. For Omaolo ESC's triage of acute cases where a nurse assessed urgent treatment as necessary, exact matches occurred in 244 out of 344 cases, representing 70.9% (CI95 65.8%, 75.7%), while matches indicating whether medical care should be sought or if self-care is sufficient occurred in 351 out of 533 cases, totaling 65.9% (CI95 61.7%, 70.0%). This is in line with the international figure of 73.3% (IQR 70.5%-82.3%) [8]. Additionally, the median app sensitivity was 51.9% (IQR 40%-78.2%), and the median specificity was 93.3% (IQR 87.4%-96.4%) [8]. Omaolo ESC exhibited a sensitivity of 62.6% and a specificity of 69.2%.

ESC's capability to detect individuals in need of immediate treatment is vital for user safety. In addition to that, an ESC that holds promise for safely assisting in self-triage and that helps preventing overcrowding of emergency rooms could bring added value to health care. Notably, concerning Omaolo, the least overtriage was observed in chief symptoms of sexually transmitted diseases (10.3%), shoulder pain stiffness or injury (8.5%), and painful or blocked ears (9.9%). Conversely, more sensitive and risk-averse chief symptom assessments such as

headaches (39.0%), heartburn (45.8%), and respiratory tract infections (47.1%) exhibited higher rates of false positives, raising concerns about overcrowding and possible unnecessary healthcare service utilization. However, due to potential serious conditions these ESCs are set to be sensitive in order to rule-out alarming symptoms.

A relatively high PPV was found in cough, dental or oral symptoms or trauma and discharge from the eye, watery or reddish eye and urinary tract infection assessments. This indicates reduced false positives, beneficial when false positives have high costs or the condition is not severe. High PPV minimizes overtreatment and unnecessary costs. Conversely, a moderate PPV found from other assessments is acceptable if follow-up tests are inexpensive and harmless precautionary measures are taken. For the symptom assessments of anal region symptoms, diarrhea, heartburn, headache and sexually transmitted diseases low number of cases has to be taken into consideration when evaluating these values.

A relatively high NPV was found in assessments of anal region symptoms, dental or oral symptoms or trauma, diarrhea, headache, heartburn, knee symptoms or injury, lower back pain or injury, sexually transmitted diseases, shoulder pain and stiffness or injury and sore throat or symptom. High NPV is crucial for serious or contagious conditions needing early intervention, minimizing false negatives. Moderate NPV observed in this study for cough, discharge from the eye, watery or reddish eye, painful or blocked ear, respiratory tract infection, and urinary tract infection is acceptable if the condition is not severe or resolves on its own.

The Matthews Correlation Coefficient shows weak positive relationships for dental or oral symptoms or trauma and lower back pain or injury, and weak negative relationships for heartburn and diarrhea. Other assessments have negligible relationships, suggesting the symptom checker's predictions are slightly better than random guessing.

Strengths and Limitations of the Study

The real-life setting presents both strengths and potential weaknesses for this study. There are notable concerns regarding potential selection bias. Firstly, users completing the ESC in Primary Healthcare (PHC) center waiting rooms may experience different symptoms compared to those using the ESC outside of such environments. Moreover, while the Omaolo ESC is designed for users over the age of 15, only individuals over the legal age of 18 were recruited for this study, potentially limiting the generalizability of findings. Furthermore, users who are unable to independently complete the ESC due to technological limitations (such as difficulty using a computer mouse or tablet devices) were excluded from the study, introducing another potential limitation.

Additionally, the rarity of users with serious acute symptoms in this setting may skew the study results, as their symptoms may not be severe enough to interfere with ESC questionnaire completion. However, it's worth noting that the Omaolo ESC prompts users to urgently contact healthcare services if they are unable to complete the questionnaire due to the severity of their symptoms.

Furthermore, potential users with mild, self-treatable symptoms may have been excluded as such cases may remain unreported in this study context. Nonetheless, the study focused on completed ESC triage assessment accuracy, specificity, and sensitivity.

In some instances, the triage classification made by nurses may have been influenced by user-reported needs for sick leave certificates, potentially biasing the calculated ESC accuracy.

The data accumulation process was hindered by the COVID-19 pandemic, which slowed down research activities [45-46]. However, the exponential growth in individual Omaolo.fi users during the pandemic, particularly with the use of the coronavirus ESC, was noted [47]. To address the scarcity of completed assessments for some ESCs and urgent cases, future research will supplement data with virtual patient cases, known as case vignettes. This approach will

also allow for the assessment of ESC performance across rare symptoms and conditions in a primary care setting.

Lastly, it's important to acknowledge that, for this study, the nurse's triage assessment was considered the gold standard. An alternative approach could have been to use outcomes during follow-up, such as re-visits or hospitalizations after 30 days, to determine success of the triage.

Implications and future research

In this study, heartburn and diarrhea were relatively infrequent as chief symptoms for safety and accuracy evaluations. However, it's crucial to monitor and supplement the safety assessments of these symptoms in future case vignette studies. Despite their limited representation in this study, the present results suggest that heartburn and diarrhea are likely safe to assess using the ESC. To address more common and less urgent situations, vignette studies will be instrumental. These vignettes, sourced from various contexts, will undergo thorough testing to ensure that individual vignettes' difficulty and correlation with overall assessment are carefully considered [48].

Moving forward, there should be an evolution toward more standardized methodologies and studies tailored to specific settings. Regulation and standardization of evaluations are vital for ensuring the transferability of findings across different contexts [49]. Additionally, adopting a patient-centered approach is essential for evaluating ESCs effectively. A standardized process with clear specifications for vignette-based clinical evaluation is necessary to guide developers and facilitate objective comparisons among ESCs. This approach will enhance the reliability and validity of ESC assessments and promote their widespread adoption in clinical practice [43, 50]

The data suggests that the Omaolo ESC is reliable for preliminary symptom assessment and triage, demonstrating a high level of safety in aligning with triage recommendations from experienced nurses, especially in acute cases. Omaolo utilizes structured questions and fixed algorithms designed by professionals to provide medically qualified recommendations, though it does not employ AI. The question of whether AI-based ESC systems would be a desirable advancement or introduce additional risk and uncertainty for patient safety is complex.

AI-based ESC systems have the potential to enhance efficiency and accuracy by continuously learning from vast datasets and adapting to evolving medical knowledge. They can rapidly process information and offer consistent assessments across different users and contexts, potentially covering a broader range of symptoms and conditions. However, there are risks associated with AI-based systems, including reliance on data quality and potential lack of nuanced clinical judgment compared to human triage nurses.

While AI-based ESC systems can complement triage processes, human supervision and oversight are essential to ensure patient safety. Human triage nurses provide contextual understanding, empathy, and critical thinking that AI systems may lack, intervening when AI-generated recommendations are uncertain or potentially harmful. Therefore, AI-based systems should be viewed as tools to augment rather than replace human triage nurses.

In conclusion, ESCs should augment traditional triage rather than substitute for it, potentially leading to benefits such as reduced phone calls and increased accessibility to health services. Omaolo ESC, with its acceptable specificity and accuracy, holds promise for preventing unnecessary use of primary healthcare services. Additionally, well-structured ESC assessments systematically collect user medical history and symptom information, evidenced by triage nurses' decisions to adjust triage based on additional user-provided information.

Impacts of the Study on the Omaolo ESC Service

The results highlighting the safety of Omaolo have been crucial for the continuation of the ESC service in Finnish health care.

Conclusions

The primary objectives of this study were to evaluate the safety and provide essential insights into the accuracy, specificity, and sensitivity of the Omaolo ESC. The results indicate that the ESC is safe for utilization compared to assessments conducted by triage nurses. Furthermore, the Omaolo ESC exhibits the potential to guide patients to appropriate triage destinations aptly, ensuring they receive timely and suitable care.

Acknowledgments

This research was carried out in addition to the work of PHC centers, municipalities, health care district–based health care organizations nationwide, Omaolo project employees, and Finnish Medical Society Duodecim employees.

Data Availability

The forms concerning individual users in the database are anonymized and cannot be combined with the user's identification code to personal data. The information in the questionnaires is coded electronically and transferred to an electronic folder. The names or personal information of the users are not shown in this electronic file but only the user's identification code. Paper forms with identification codes are stored in a locked cabinet. The consent form containing the user's name and signature is kept in a separate locked cabinet. All paper forms are stored in the Tampere University. The forms are disposed of with a shredder 10 years after the end of the study. The electronic data sets generated during and analyzed in this study are available from the corresponding author on reasonable request.

Conflicts of Interest

Minna Kaila (MD, PhD, professor) is the president of the Finnish Medical Society Duodecim.
Tuomas H. Koskela worked as a salaried part-time editor for Duodecim Medical publications until 2018.
Ville Liu works in the MD PhD program of the Faculty of Medicine at the University of Helsinki.

Abbreviations

GP: General practitioner
ESC: Electronic symptom checker
PHC: Primary health care
MSAH: Finnish Ministry Of Social Affairs
FIHW: Finnish Institute for Health and Welfare

Multimedia Appendix 1:

Quantitative clinical study

Multimedia Appendix 2:

Details of affirmative answers to "do you feel the need to change your assessment of need for treatment (triage) after seeing the responses and recommendation from the electronic symptom checker?".

Multimedia Appendix 3:

Identified conflict situations (21 critical assesments). The findings of the inspection, the identified conflict, the final result, and the declared final result (assessed by the evaluation

References

1. Hochberg I, Allon R, Yom-Tov E. Assessment of the Frequency of Online Searches for Symptoms Before Diagnosis: Analysis of Archival Data. J Med Internet Res. 2020;22(3):e15065. Published 2020 Mar 6. doi:10.2196/15065
2. Sebastian Cross, Ahmed Mourad, Guido Zuccon, and Bevan Koopman. 2021. Search Engines vs. Symptom Checkers: A Comparison of their Effectiveness for Online Health Advice. In Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 11 pages. Doi: 10.1145/3442381.3450140

3. EPatient survey 2020. Health & Care Management. 2020. URL: https://wdww.hcm-magazin.de/epatient-survey-2020-digital-health-studies/150/10992/407743 [accessed 2021-03-06]

4. Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study JMIR Public Health Surveill 2021;7(1):e22637 doi: 10.2196/22637
5. Tasnim S, Hossain MM, Mazumder H. Impact of Rumors and Misinformation on COVID-19 in Social Media. J Prev Med Public Health. 2020;53(3):171-174. doi:10.3961/jpmph.20.094
6. McLellan J, Heneghan C, Roberts N, Pluddemann A. Accuracy of self-diagnosis in conditions commonly managed in primary care: diagnostic accuracy systematic review and meta-analysis. BMJ Open. 2023 Jan 10;13(1):e065748. doi: 10.1136/bmjopen-2022-065748.
7. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med. 2020 Jan 13;3:4. doi: 10.1038/s41746-019-0206-x.
8. Schmieding M, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel M Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation J Med Internet Res 2022;24(5):e31810 DOI: 10.2196/31810
9. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. J Pers Med. 2020;10(2):21. Published 2020 Mar 31. pmid:32244292

10. Morse KE, Ostberg NP, Jones VG, Chan AS 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(11):e20549 doi: 10.2196/20549
11. van Veen T, Binz S, Muminovic M, Chaudhry K, Rose K, Calo S, Rammal JA, France J, Miller JB. Potential of Mobile Health Technology to Reduce Health Disparities in Underserved Communities. West J Emerg Med. 2019 Aug 6;20(5):799-802. doi: 10.5811/westjem.2019.6.41911

12. Rodgers M, Raine G, Thomas S, Harden M, Eastwood A. Informing NHS policy in 'digital-first primary care': a rapid evidence synthesis. Southampton (UK): NIHR Journals Library; 2019 Dec. PMID: 31869020. doi: 10.3310/hsdr07410
13. Evaluation of Babylon GP at hand Final evaluation report Prepared by Ipsos MORI and York Health Economics Consortium with Prof. Chris Salisbury for NHS Hammersmith and Fulham CCG and NHS England, 2019 http://allcatsrgrey.org.uk/wp/download/informatics/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf
14. NHS 111 Minimum Data Set, England, March 2021 https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2021/04/NHS-111-MDS-Apr-2021-Statistical-Note.pdf
15. Aboueid S, Meyer SB, Wallace JR, Mahajan S, Nur T, Chaurasia A. Use of symptom checkers for COVID-19-related symptoms among university students: a qualitative study. BMJ Innov. 2021 Apr;7(2):253-260. doi: 10.1136/bmjinnov-2020-000498
16. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, Moriarty T, Gleason N, Intinarelli G, Gonzales R. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc. 2020 Jun 1;27(6):860-866. doi: 10.1093/jamia/ocaa051.
17. Laukka E, Kujala S, Gluschkoff K, Kanste O, Hörhammer I, Heponiemi T. Leaders' support for using online symptom checkers in Finnish primary care: Survey study. Health Informatics Journal. 2021;27(4). doi:10.1177/14604582211052259
18. Kujala S, Hörhammer I Health Care Professionals' Experiences of Web-Based Symptom Checkers for Triage: Cross-sectional Survey Study J Med Internet Res 2022;24(5):e33505 doi: 10.2196/33505
19. Thompson, C. & Dowding, C. (2009) Essential Decision Making and Clinical Judgement for Nurses.
20. National Academies of Sciences, Engineering, and Medicine (2015). Balogh, Erin P; Miller, Bryan T; Ball, John R

(eds.). Improving Diagnosis in Health Care. Washington, DC: The National Academies Press. p. S-1. doi:10.17226/21794.
21. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, et al. (2021) Accuracy of online symptom checkers and the potential impact on service utilisation. PLOS ONE 16(7): e0254088. doi:10.1371/journal.pone.0254088
22. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study BMJ 2015;351:h3480
23. G Gilbert, Stephen et al. "How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs." BMJ open vol. 10,12 e040269. 16 Dec. 2020, doi: 10.1136/bmjopen-2020-040269
24. Semigran, Hannah L et al. "Comparison of Physician and Computer Diagnostic Accuracy." JAMA internal medicine vol. 176,12 (2016): 1860-1861. doi:10.1001/jamainternmed.2016.6001
25. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a Popular Online Symptom Checker for Ophthalmic Diagnoses. JAMA Ophthalmol. 2019;137(6):690–692. doi:10.1001/jamaophthalmol.2019.0571
26. Munsch N, Martin A, Gruarin S, et al. Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study. J Med Internet Res. 2020;22(10):e21299. Published 2020 Oct 6. doi:10.2196/21299
27. Wallace, W., Chan, C., Chidambaram, S. et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. npj Digit. Med. 5, 118 (2022).
28. Yuya Yoshida, Glenn Thomas Clark, Accuracy of online symptom checkers for diagnosis of orofacial pain and oral medicine disease, Journal of Prosthodontic Research, 2021, Volume 65, Issue 2, Pages 186-190, Released on J-STAGE June 30, 2021, Advance online publication September 09, 2020, Online ISSN 1883-9207, Print ISSN 1883-1958, https://doi.org/10.2186/jpr.JPOR_2019_499,
29. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S-C, Leung L-P. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. Hong Kong Journal of Emergency Medicine. 2020;27(4):217-222. doi:10.1177/1024907919842486
30. El-Osta A, Webber I, Alaa A, et al What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study BMJ Open 2022;12:e053566. doi: 10.1136/bmjopen-2021-053566

31. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient Perspectives on the Usefulness of an Artificial Intelligence-Assisted Symptom Checker: Cross-Sectional Survey Study. J Med Internet Res. 2020;22(1):e14679. Published 2020 Jan 30. doi: 10.2196/14679
32. Liu, V. D., Sellgren, L., Kaila, M., & Koskela, T. (2021). Usability of online symptom checkers. Finnish Journal of EHealth and EWelfare, 13(2), 100–112. https://doi.org/10.23996/fjhw.97020
33. Baker A, Perov Y, Middleton K, et al. A Comparison of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis. Front Artif Intell. 2020;3:543405. Published 2020 Nov 30. doi:10.3389/frai.2020.543405
34. Hill, M.G., Sim, M. and Mills, B. (2020), The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med. J. Aust., 212: 514-519. https://doi.org/10.5694/mja2.50600
35. Fraser HSF, Cohan G, Koehler C, Anderson J, Lawrence A, Pateña J, Bacher I, Ranney ML Evaluation of Diagnostic and Triage Accuracy and Usability of a Symptom Checker in an Emergency Department: Observational Study JMIR Mhealth Uhealth 2022;10(9):e38364 doi: 10.2196/38364PMID: 36121688PMCID: 9531004
36. Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS Health Information Seeking From an Intelligent Web-Based Symptom Checker: Cross-sectional Questionnaire Study J Med Internet Res 2022;24(8):e36322 doi: 10.2196/36322PMID: 35984690PMCID: 9440406
37. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A Online Symptom Checkers: Recommendations for a Vignette-Based Clinical Evaluation Standard J Med Internet Res 2022;24(10):e37408 doi: 10.2196/37408PMID: 36287594PMCID: 9647454

38. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar doi: 10.2196/24475

39. https://digifinland.fi/wp-content/uploads/2023/07/FI22_871876_Iss4_DigiFinland_Oy_EU-QMS_2023-05-31.pdf
40. Liu VDM, Kaila M, Koskela T, User initiated symptom assessment with an electronic symptom checker. Study protocol for mixed-methods validation. JMIR Research Protocol DOI: 10.2196/41423
41. Omaolo instructions for use: https://www.omaolo.fi/kayttoohjeet/omaolo-instructions.pdf DigiFinland Oy (owner: Republic of Finland), Kuntatalo, Toinen linja 14, 00530 Helsinki.

42. https://digifinland.fi/omaolohelp/verkkokoulutus/2-oirearviot/2-1-oirearvio/prioriteetti-eli-kiireellisyys/

43. Singh H, Meyer AN, Thomas EJ, The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014; 23: 727-731

44. Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, Wicks P, Wolanski J, Tutschkow D, Schade Brittinger C, Timmermann L, Jerrentrup A Safety of Triage Self-assessment Using a Symptom Assessment App for Walk-in Patients in the Emergency Care Setting: Observational Prospective Cross-sectional Study JMIR Mhealth Uhealth 2022;10(3):e32340 doi: 10.2196/32340

45. Valmorri, L., Vertogen, B., Zingaretti, C. et al. Clinical research activities during COVID-19: the point of view of a promoter of academic clinical trials. BMC Med Res Methodol 21, 91 (2021). https://doi.org/10.1186/s12874-021-01291-0

46. Nguyen HL, Tran OT, Ha DA, Phan VH, Nguyen CT, et al. (2021) Impact of the COVID-19 pandemic on clinical research activities: Survey of study participants and health care workers participating in a hypertension trial in Vietnam. PLOS ONE 16(7): e0253664. https://doi.org/10.1371/journal.pone.0253664

47. https://digifinland.fi/omaolo-palvelu-laajenee-taman-vuoden-aikana-kattamaan-lahes-koko-suomen/

48. How suitable are clinical vignettes for the evaluation of symptom checker apps? A test theoretical perspective Marvin Kopka, Markus A. Feufel, Eta S. Berner, Malte L. Schmiedin medRxiv 2023.05.23.23290410; doi: https://doi.org/10.1101/2023.05.23.23290410

49. Pairon Anthony, Philips Hilde, Verhoeven Veronique, A scoping review on the use and usefulness of online symptom checkers and triage systems: How to proceed? Frontiers in Medicine9,2023 DOI:10.3389/fmed.2022.1040926

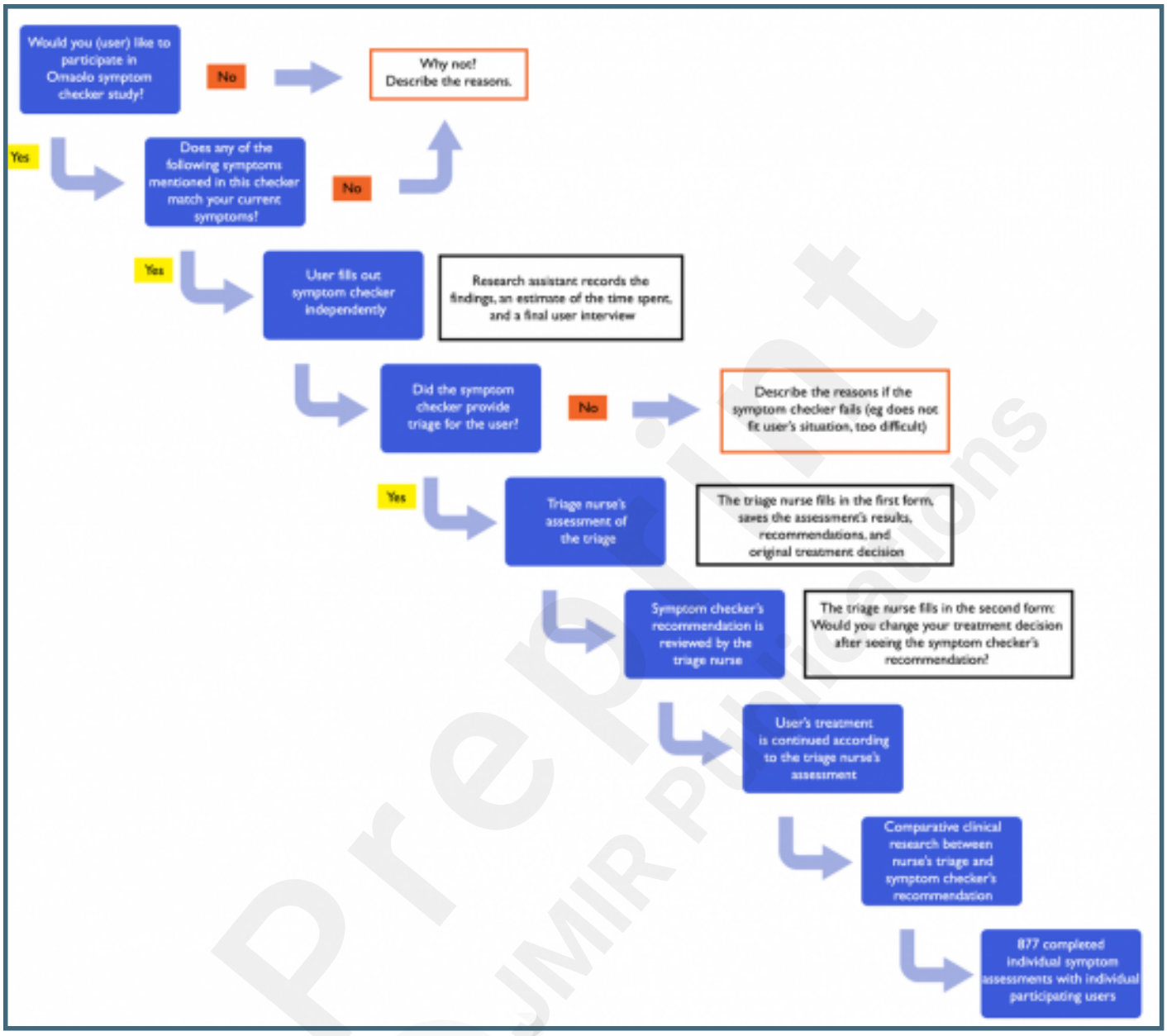
50. Riboli-Sasco E, El-Osta A, Alaa A, Webber I, Karki M, El Asmar ML, Purohit K, Painter A, Hayhoe B Triage and Diagnostic Accuracy of Online Symptom Checkers: Systematic Review J Med Internet Res 2023;25:e43803 doi: 10.2196/43803PMID: 37266983PMCID: 10276326

Supplementary Files

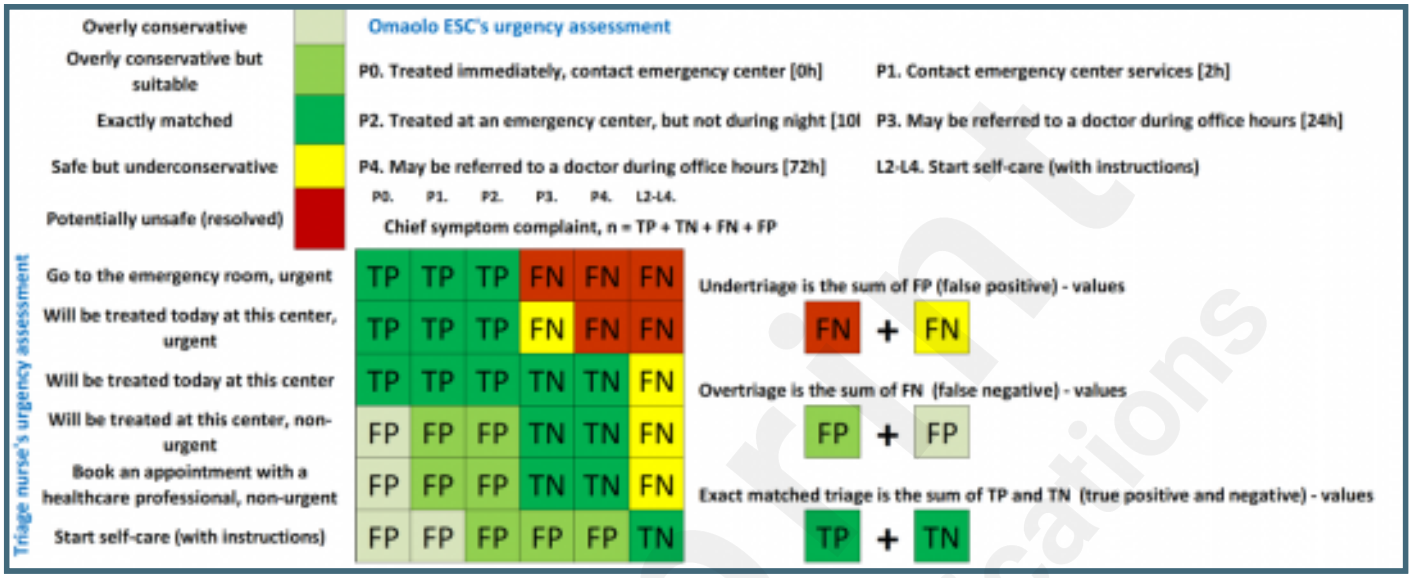
Untitled.
URL: http://asset.jmir.pub/assets/1098ac4ea19808582e80c2f8b5ccbf14.xlsx

Figures

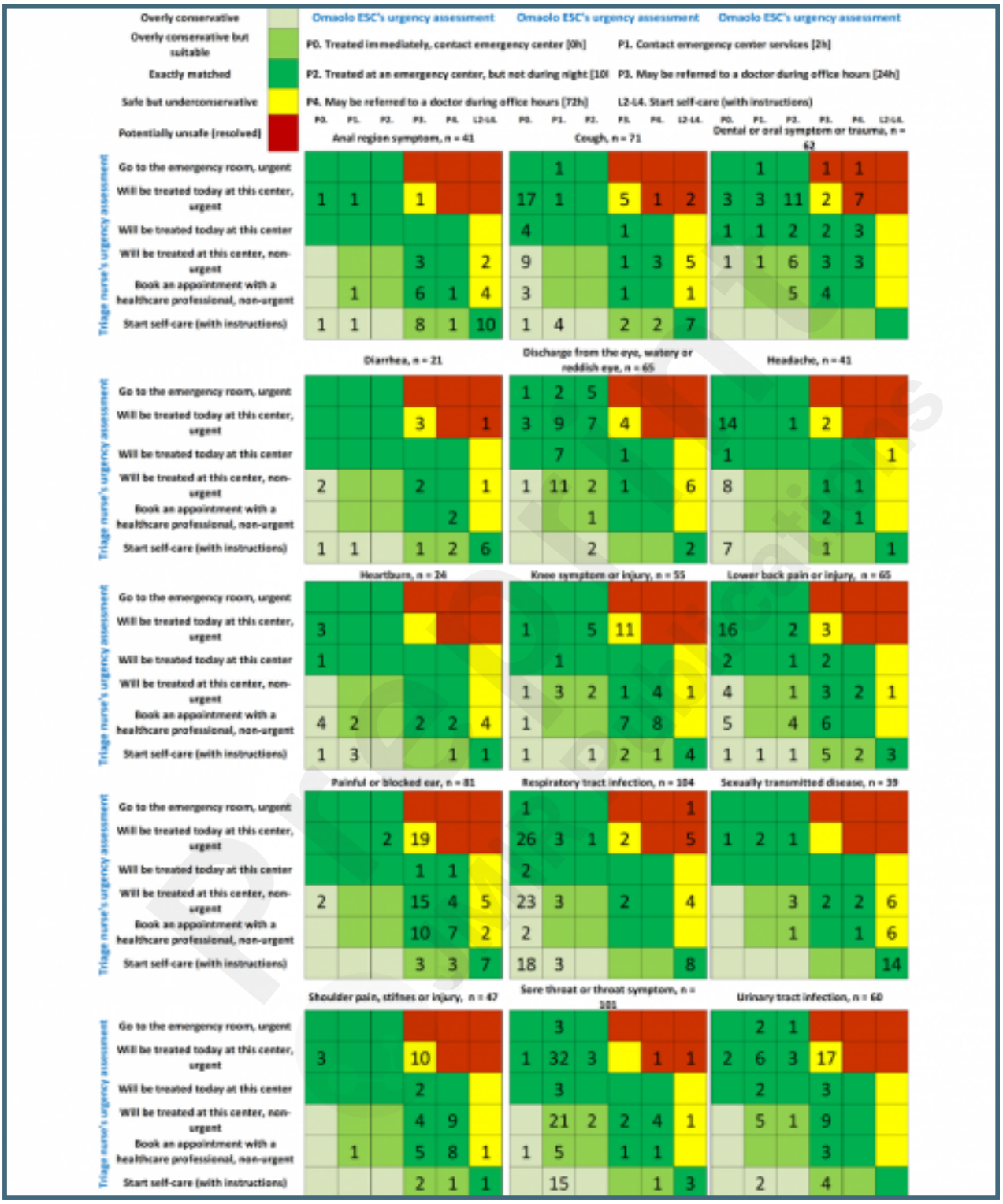
Flow chart showing the collection of the symptom checker and nurse triage data.



An example showing how the triage comparison chart (confusion matrices) in Figure 3 were constructed with color-coded differences in triage levels (overly conservative, overly conservative but suitable, exactly matched, safe but under conservative and potentially unsafe (resolved)). Matching rows (triage nurse) to their respective columns (symptom assessment) results in a safety assessment. P0-P4 = classification of emergency care criteria. L2-4 = referral urgency classes. The columns show how often, by ESC, the symptom assessment was overestimated, underestimated, or accurately matched compared to the assessment made by the triage nurse and the decision made on further referral.



Triage comparison chart (confusion matrices) with color-coded differences in triage levels.



The individual bar heights (Y axis, %) reflect the proportional correspondence and accuracy levels of the ESC triage with the nurse's triage (overly conservative, overly conservative but suitable, exactly matched, safe but under conservative and potentially unsafe (resolved)). Individual ESCs are depicted on the X axis.

Multimedia Appendixes

Quantitative clinical study.
URL: http://asset.jmir.pub/assets/edd795fde1375e8d762ed33726ed1fde.docx

Details of affirmative answers to “do you feel the need to change your assessment of need for treatment (triage) after seeing the responses and recommendation from the electronic symptom checker?".
URL: http://asset.jmir.pub/assets/643551892672533d8421bcf08e7b717f.docx

Identified conflict situations (21 critical cases). The findings of the inspection, the identified conflict, the final result, and the declared final result (assessed by the evaluation team).
URL: http://asset.jmir.pub/assets/9bfd9cdb2634fff428a75daf6478ae1e.docx