# Safety of patient self-triage: real-life prospective evaluation of a symptom-checker in adult patients visiting an interdisciplinary emergency care center

Andreas Meer, Philipp Rahm, Markus Schwendinger, Michael Vock, Bettina Grunder, Jacopo Demurtas, Jonas Rutishauser

Submitted to: Journal of Medical Internet Research
on: March 07, 2024

# *Table of Contents*

# Safety of patient self-triage: real-life prospective evaluation of a symptom-checker in adult patients visiting an interdisciplinary emergency care center

Andreas Meer[1] MD, MSc; Philipp Rahm[2] MD; Markus Schwendinger[2] MD; Michael Vock[3] PhD; Bettina Grunder[1] MD; Jacopo Demurtas[1] MD, PhD; Jonas Rutishauser[4] MD

[1]In4medicine Inc Bern CH
[2]Emergency Department Cantonal Hospital Baden Baden CH
[3]Institute of Mathematical Statistics and Actuarial Science University of Berne Berne CH
[4]Clinical trial unit Cantonal Hospital Baden Baden CH

**Corresponding Author:**
Andreas Meer MD, MSc
In4medicine Inc
Monbijoustrasse 23
Bern
CH



## *Abstract*

**Background:** Symptom-checker self-triage apps assist patients to determine the urgency of medical care. To be safe and effective, these tools must be validated, particularly to avoid potential hazardous undertriage.

**Objective:** To investigate the safety of patient's self-triage using a symptom-checker.

**Methods:** A single centre, prospective clinical trial comparing the individual outcomes of patients' self-triage with the assessment of the clinical urgency made by three successive interdisciplinary panels of physicians. Data collected between 25 November 2019 and 1 May 2020. Panel assessments and data analysis completed on 29 August 2022.
Setting: Walk-in-Clinic and Interdisciplinary Emergency Department (WIC/ED) of the cantonal hospital of Baden, Switzerland.
Partcipants: All patients ? 18 years attending the WIC/ED between 8:00 a.m. and 5:00 p.m. Exclusion criteria included ESI 1 and presence of symptoms not encompassed by the symptom-checker.
22 676 (12 655 ER, 10 021 WIC) used the facility during the enrollment period. 7550 patients attended WIC/ED within the recruitment window. 2550 gave informed consent, 7 patients withdrew.
Intervention: Participants assessed their symptoms using SMASSPathfinder, a web-based symptom-checker based on a computerized transparent neural network.
Main Outcome and Measure: The assessment by the panels encompassed the appropriate time-to-treat and the adequate point-of-care. If a case was adjudicated as undertriaged by the first two panels, the third panel assessed the patient's risk to health or life, making a decision on whether a potentially hazardous undertriage had been present. Using a Clopper-Pearson confidence interval, we assumed that in order to confirm the symptom-checkers safety, the upper confidence bound should lie below 1%.

**Results:** 2543 patients were included in the study, 1227 (48.25%) female, 1316 (51.75%) male, 1397 (54.94) 18-49 y, 668 (26.27%) 50.65y, 360 (14.16%) 66-80y, 118 (4.64%) >80y. Of the 2543 cases none reached the pre-specified criterion for a potentially hazardous undertriage. This resulted in an upper 95% confidence bound for the probability of a potentially hazardous undertriage of 0.1184%.

**Conclusions:** The symptom checker proved to be a safe triage tool, avoiding undertriage in a real-life clinical setting of emergency consultations at a WIC/ED. Our data suggest the symptom checker may be safely used in clinical routine. Clinical Trial: ClinicalTrials.gov (NCT04055298)

(JMIR Preprints 07/03/2024:58157)
DOI: https://doi.org/10.2196/preprints.58157

### 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

**Authors' information/metadata**

Andreas Meer[a] MD, Philipp Rahm[b] MD, Markus Schwendinger[b] MD, Michael Vock[c] PhD, Bettina Grunder[a] MD, Jacopo Demurtas[a] MD, Jonas Rutishauser[d] MD

[a]In4medicine, Inc., Bern, Switzerland

[b]Cantonal Hospital Baden, Emergency Department, Baden, Switzerland

[c]Institute for Mathematical Statistics and Actuarial Science, University of Bern, Switzerland

[d]Cantonal Hospital Baden, Clinical Trial Unit, Baden, Switzerland

Corresponding Author:

Andreas Meer, MD, MSc

Monbijoustrasse 23

CH-3011 Bern

Switzerland

E-Mail: a.meer@in4medicine.ch

**Type of Paper**

Original Paper

**Length of Paper**

Word count: 3300

# Safety of patient self-triage: real-life prospective evaluation of a symptom-checker in adult patients visiting an interdisciplinary emergency care center

**Key words**

safety, telemedicine, teletriage, symptom-checker, self-triage, self-assessment.

**Trial registration**

ClinicalTrials.gov (NCT04055298)

**Abstract**

Background

Symptom-checkers have become important tools for self-triage, assisting patients to determine the urgency of medical care. To be safe and effective, these tools must be validated, particularly to avoid potential hazardous undertriage without leading to inefficient overtriage. Only limited safety data from studies including small sample sizes have been available so far.

Objective

The objective of our study was to prospectively investigate the safety of patients' self-triage in a large patient sample. We used SMASS pathfinder, a symptom-checker based on a computerized transparent neural network.

Methods:

We recruited 2543 patients into this single centre, prospective clinical trial conducted at the cantonal hospital of Baden, Switzerland. Patients with an Emergency Severity Index of 1-2 were treated by the team of the emergency department, while those with an index of 3-5 were seen at the walk-in clinic by general physicians. We compared the triage recommendation obtained by the patients' self-triage with the assessment of the clinical urgency made by three successive interdisciplinary panels of physicians (Panel A, B, C). Using a Clopper-Pearson confidence interval, we assumed that in order to confirm the symptom-checkers safety, the upper confidence bound for the probability of a potentially hazardous undertriage should lie below 1%. A potentially hazardous undertriage was defined as a triage in which either all (consensus criterion) or the majority (majority criterion) of the experts of the last panel (Panel C) rated the triage of the symptom-checker to be "rather likely" or "likely" life-threatening or harmful.

Results:

Of the 2543 patients, 1227 (48.3%) were female and 1316 (51.7%) male. None of the patients reached the pre-specified consensus criterion for a potentially hazardous undertriage. This resulted in an upper 95% confidence bound of 0.1184%. 4 cases met the majority criterion. This resulted in an upper 95% confidence bound for the probability of a potentially hazardous undertriage of 0.3616%. The two-sided 95% Clopper-Pearson confidence interval for the probability of overtriage (450 cases, 17.7%) was 16.23% to 19.24%, which is considerably lower than figures reported in the literature.

Conclusions:

The symptom-checker proved to be a safe triage tool, avoiding potentially hazardous undertriage in a real-life clinical setting of emergency consultations at a WIC/ED, whithout causing undesirable overtriage. Our data suggest the symptom-checker may be safely used in clinical routine.

Study Registration:

ClinicalTrials.gov identifier NCT04055298

**Introduction**

In potentially critical situations, clinical warning signs and symptoms may be considered too late by patients due to lack of professional triage [1]. In this context, various initiatives have been launched to improve outpatient emergency care and the population's access to a low-threshold initial medical assessment [2]. Symptom-checkers, which enable medical self-triage, have recently been introduced for this purpose. Such tools could assist the increasing number of persons without ready access to a primary care physician, e.g. migrants or young persons who had been previously healthy. If implemented in settings outside the hospital, i.e. at home or at work, tools for efficient and safe self-triage could help avoid unnecessary emergency hospital visits, thus contributing to reduce overcrowding and costs.

In order to fulfil the regulatory requirements [3] and to be used as a part of standard care, the appropriateness and safety of these instruments must be evaluated in concrete clinical settings with real patients [4] [5] [6].

Appropriate care results from adequate triage and treatment, while inappropriate care may lead to unsuitable or even dangerous health care delivery. The concept of appropriateness hence includes a widespread range of quality aspects, of which safety is only one. The difficulty of assessing appropriateness in health care and of gaining agreement between clinicians on acceptable and safe care is highlighted by different authors [7] [8]. When assessing the appropriateness of medical triage, the question “was the decision right” suggests that there is merely one single correct triage decision. This question does not appropriately reflect the complex interaction of clinical, social and environmental factors in medical decision-making. Rather, physicians should consider a range of appropriate triage decisions to guide their actions. Safety is an essential quality attribute of a medical service. In contrast to the idea of appropriateness, the concept of safety focuses on the risk of a specific conduct. When asking about the safety of a symptom-checker, a risk-based approach should be taken, and safety should encompass possible risks for patients’ health and life [3].

An evaluation of 23 symptom-checkers using 45 patient vignettes concluded that most symptom-checkers were deficient in both appropriate triage and correct diagnosis [9]. However, the study did not comment on the safety of the tested devices. A review article including 14 studies found inconsistent evidence regarding the triage and diagnostic appropriateness of symptom-checkers for common health problems. The average appropriateness of triage ranged from 27% to 92%. This article did not specifically evaluate the safety of symptom-checkers [5]. Another review article cited only six studies that analysed the safety of symptom-

checkers [10]. These studies were mostly short-term and included samples that were too small and heterogeneous to make reliable statements about safety.

In view of the present shortage of data on self-triage, we aimed to investigate the safety of a newly developed symptom-checker (SMASS: Swiss Medical Assessment System) in a concrete clinical setting with patients seeking emergency care.

**Methods**

Prior to inclusion of the first patient, the study was registered (ClinicalTrials.gov identifier: NCT04055298). The study was performed between November 25, 2019 and May 1, 2020, at the Walk-in-Clinic and Interdisciplinary Emergency Department (WIC/ED) of the cantonal hospital of Baden, Switzerland. The WIC/ED is open 24 hours a day for 365 days a year and treats about 55 000 patients annually. Patients are routinely triaged by a nurse using the Emergency Severity Index (ESI) [11]. ESI 1-2 patients are treated in the Emergency Department, while ESI 3-5 patients are treated in the Walk-in-Clinic.

The symptom-checker used in the study (SMASS in the Pathfinder version, Release 4.1.12), was developed by In4medicine, Inc. The first author (author AM) is the CEO and founder of this company. To minimize bias, a majority of independent researchers were involved in the study, including establishing the protocol and all practical aspects of the trial. No employee of In4medicine took part in the actual conduct of the trial. Data analysis and the statistical calculations were performed by an independent biostatistician. The study was independently monitored by the clinical trial unit (CTU) of the Medical Faculty of the University of Bern.

The SMASS pathfinder symptom-checker is a Medical Device Class I under the Medical Device Directive (MDD) and Medical Device Class IIb under the Medical Device Regulation (MDR). The CE declaration of conformity to the Swiss Agency for Therapeutic Products (Swissmedic) was made on 4 June, 2018. The symptom-checker is a web-based software that aims to support health professionals and lay persons in the structured documentation and assessment of health problems and to advise users about possible medical assessment steps and treatment measures. It is based on a computerized neural network that incorporates extensive data from scientific studies, guidelines, and expertise from various professional boards of specialists in the field of prehospital medical triage. The symptom-checker provides digitalised questionnaires of 125 frequent reasons for consultations (e.g. fever, cough, abdominal pain) and their associated red flags. Based on the triage result, a report including patient gender, age group, symptoms, medical history and recommendations as to the appropriate time-to-treat (T2T) and point-of-care (PoC) is provided. Depending on the presence of red flags, the symptom-checker assigns the clinical condition of the patient to a triage level (Tables 1 and 2). If five or more assessment questions are answered as "unclear," the user is notified that the software cannot provide targeted triage advice and that the patient should seek immediate consultation with a physician concerning his or her medical complaints.

| | AMBULANCE | HOSPITAL | DOCTOR | CALLCENTER | PHARMACY | SELFCARE | UNCLEAR |
|---|---|---|---|---|---|---|---|
| EMERGENCY | 16 | 15 | | | | | |
| IMMEDIATELY | 14 | 13 | 12 | | | | |
| TODAY | | 11 | 10 | 8 | 6 | 4 | |
| LATER | | | 9 | 7 | 5 | 3 | 1 |
| UNCLEAR | | | | | | 2 | 0 |

Tab 1. Triage Levels as recommended to the patient by the symptom-checker. Recommendations are given regarding time-to-treat (Emergency, Immediately, Today, Later, Unclear) and Point-of-Care (Ambulance, Hospital, Physician, Call Center, Pharmacy, Selfcare, Unclear)

| Triage Level | Name | Recommended action |
|---|---|---|
| 16 | Emergency Ambulance | CPR/CPR readiness. There is a potentially life-threatening condition. Medical treatment must be given now. Alert the emergency services via the number 144. |
| 15 | Emergency Hospital | CPR/CPR readiness. There is a potentially life-threatening condition. Medical treatment must be given now. Alert the emergency services via the number 144. Medical treatment should be provided at a hospital. |
| 14 | Immediately Ambulance | Medical treatment does not allow any delay. Treatment should be given immediately. Alert the emergency services via the number 144. |
| 13 | Immediatel Hospital | Medical treatment does not allow any delay. Treatment should be given immediately. Medical treatment should be provided at a hospital. |
| 12 | Immediately Doctor | Medical treatment does not allow any delay. Treatment should be given immediately. Medical treatment should be provided by a registered doctor [1] |
| 11 | Today Hospital | Medical treatment does not have to take place immediately, but should not be delayed until tomorrow or over the weekend. Medical treatment should take place within the next 24 hours. Medical treatment should be provided at a hospital. |
| 10 | Today Doctor | Medical treatment does not have to take place immediately, but should not be delayed until tomorrow or over the weekend. Medical treatment should take place within the next 24 hours. Medical treatment should be provided by a registered doctor [1] |
| 9 | Later Doctor | Medical treatment is not urgent. If the symptoms do not subside in the next 2 days, treatment by a doctor is indicated. Medical |

1 e.g. family doctor, family doctor substitute, family doctor emergency service or suitable specialist

| | | |
|---|---|---|
| | | treatment should be provided by a registered doctor [1] |
| 8 | Today Call Centre | Medical treatment does not have to take place immediately, but should not be delayed until tomorrow or over the weekend. Medical treatment should take place within the next 24 hours. The affected person should be advised by a telemedicine centre on how to proceed. |
| 7 | Later call centre | Medical treatment is not urgent. If the symptoms do not subside in the next 2 days, treatment by a doctor is indicated. The affected person should be advised by a telemedicine centre on how to proceed. |
| 6 | Today Pharmacy | Medical treatment does not have to take place immediately, but should not be delayed until tomorrow or over the weekend. Medical treatment should take place within the next 24 hours. The affected person should be advised at a pharmacy on how to proceed. |
| 5 | Later Pharmacy | Medical treatment is not urgent. If the symptoms do not subside in the next 2 days, treatment by a doctor is indicated. The affected person should be advised at a pharmacy how to proceed. |
| 4 | Today Selfcare | Medical treatment does not have to take place immediately, but should not be delayed until tomorrow or over the weekend. Medical treatment should take place within the next 24 hours. The complaints can be treated independently by simple measures. |
| 3 | Later Selfcare | Medical treatment is not urgent. If the symptoms do not subside in the next 2 days, treatment by a doctor is indicated. The complaints can be treated independently by simple measures. |
| 0-2 | Unclear | The survey contains too many ambiguities. A targeted initial assessment is not possible. |

Tab. 2. Recommended actions to be taken by the patient, as defined by triage levels (left column). Levels range from 0 (lowest level) to 16 (highest level). Interpretations of the triage level and measures to be taken are specified in the right column.

All patients ≥ 18 years attending the WIC/ED between 8:00 a.m. and 5:00 p.m. were eligible. Exclusion criteria included age <18-years; ESI 1 patients requiring immediate, life-saving intervention; inability to use a tablet PC; inability to communicate in German, French, Italian or English; inability or unwillingness to give written informed consent and follow the procedures of the study; known or suspected non-compliance; known drug or alcohol abuse; presence of symptoms or complaints not encompassed by the symptom-checker data base (e.g. long-lasting hiccups, hair loss).

After instruction by the study staff and providing written informed consent, the participants independently assessed their health status and complaints as instructed by the symptom-checker on a tablet PC. They were subsequently evaluated and treated by routine medical staff.

In Primary Care, medical triage decisions usually have to be based solely on the patient's symptoms. We have chosen experts independent of the treatment (Panel A, B, C) as evaluators to ensure that the triage decision is based purely on the symptoms of the study patients. Including treating physicians as comparators

in the study could have influenced the triage decision by additional information (physical examination, diagnostic test results).

Our evaluation of the symptom-checker focused on safety, as this is an essential quality attribute of a medical device [3]. In order to reflect the highly individual nature of medical decision making, which usually results in low inter-rater reliability [12] [13] [14], an independent team of experienced physicians engaged in a stepwise evaluation procedure, in which each case that was classified as undertriaged by the Panel A experts, was assessed by several experts.

A research assistant and three external interdisciplinary panels of board-certified physicians were involved in the evaluation process (Panel A, five experts; Panel B, two experts; Panel C, five experts). Except for one of the 12 panellists (author BG), they were not affiliated with In4Medicine, Inc. None of them took part in the conduct of the study. For every patient, the symptom-checker issued a report summarizing the clinical information. Patients and panellists were unaware of the triage level recommendations (T2T and PoC) made by the symptom-checker. All reports were first assessed by members of Panel A, who adjudicated an appropriate range of triage levels to every case. The research assistant then compared the adjudication of the Panel A experts with the recommendation issued by the symptom-checker. If the comparison showed that the recommendation of the symptom-checker was below the appropriate range of triage levels determined by the rater of panel A, hence was undertriaged, the case was assigned to panel B. In 80 instances, panellists erroneously examined the same cases twice and concluded on diverse triage recommendations. In these cases, the first of the two recommendations were used for the analysis.

The evaluation procedure was repeated by Panel B. Each of the two panellists evaluated all diverging cases. If the case was undertriaged according to three experts (one expert from Panel A, two experts from Panel B), the case was subsequently analysed by Panel C.

Each member of Panel C individually assessed the clinical safety of the triage decision based on the complete structured reports generated by the symptom-checker as well as the WIC/ED's redacted discharge reports. Each of the five panellists decided individually on potentially hazardous undertriage. In a modified Delphi process, the panellists first individually adjudicated potentially hazardous undertriage on a 4-Point Likert-Scale. Possible ratings were "unlikely", "rather unlikely", "rather likely" and "likely" that the patient was exposed to a risk to life or health. If the panellists subsequently reached a consensus that the triage of

the symptom-checker was “rather likely” or “likely” exposing a patient to a risk to life or health, the case was considered a potentially hazardous undertriage (consensus criterion). As a complement to the original analysis plan, a modified criterion for potentially hazardous undertriage was evaluated, defined as a majority of Panel C members judging a risk to life or health as “rather likely” or “likely” (majority criterion).

The primary analysis consisted of the calculation of a 95% upper Clopper-Pearson confidence bound for the probability of undertriage resulting in a risk to life or health (potentially hazardous undertriage). In order to confirm the safety of the symptom-checker, this upper confidence bound should lie below 1%. For the sample size calculation, we assumed that a 20% probability of failure to meet this criterion is acceptable for a true probability of potentially hazardous undertriage of no more than 0.5%. This is equivalent to requiring a one-sided test at level 5% to show that the probability of potentially hazardous under-triage is below 1% with a power of 80%, assuming that the true probability is 0.5%. This resulted in a minimal sample size of 2185 patients. Accounting for an estimated rate of 2% “unclear” responses, at least 2230 patients were planned to be included. Secondary analyses included central 95% Clopper-Pearson confidence intervals for the further probabilities, based on corresponding empirical proportions. The software R version 4.2.0 (Windows, 64-bit) was used for the statistical evaluations.

Ethical considerations:

The study was approved by the competent ethics committee (Ethikkommission Nordwest- und Zentralschweiz EKNZ, project ID 01784) and was conducted in accordance with the most recent version of the Declaration of Helsinki, complying with ICH-GCP and ISO EN 14155 (clinical investigation of medical devices for human subjects – good clinical practice) as well as with applying national legal and regulatory requirements. All patients gave written informed consent to participate in the study. They did not receive any financial or other compensation. Patients were anonymized upon data collection. Discharge notes studied by panel C were redacted,

Generative artificial intelligence was not used in any portion of the manuscript writing.

**Results**

Baseline characteristics of the participants are shown in Table 3, and the recommendations obtained by the symptom-checker in Table 4. Figure 1 shows the flow of analyses by Panels A – C.

| Characteristics | | Cases |
|---|---|---|
| **Age (years), n (%)** | | |
| | 18-49 | 1397 (54.9) |
| | 50-65 | 668 (26.3) |
| | 66-80 | 360 (14.2) |
| | >80 | 118 (4.6) |
| **Gender, n (%)** | | |
| | Female | 1227 (48.3) |
| | Male | 1316 (51.7) |
| **Reason for Encounter, n (%)** (15 most frequent) | | |
| | Stomach pain | 287 (11.3) |
| | Chest pain | 168 (6.6) |
| | Lumbar back pain | 144 (5.7) |
| | Urinary tract problems | 124 (4.9) |
| | Trauma/Fall | 121 (4.8) |
| | Headache | 90 (3.5) |
| | Dizziness | 87 (3.4) |
| | Wound/skin injury | 82 (3.2) |
| | Foot injury (caused by an accident) | 81 (3.2) |
| | Leg problems | 74 (2.9) |
| | Breathlessness | 69 (2.7) |
| | Cold/influenza infection | 64 (2.5) |
| | Finger injury (caused by an accident) | 55 (2.2) |
| | Knee injury (caused by an accident) | 51 (2.0) |
| | Hand injury (caused by an accident) | 43 (1.7) |

Tab. 3. Characteristics of the study population (N = 2543)

| **Triage Level, n (%)** | Cases |
|---|---|
| Emergency Ambulance | 57 (0.0224) |
| Emergency Hospital | 142 (0.0558) |
| Immediately Ambulance | 2 (0.0008) |
| Immediately Hospital | 685 (0.2694) |
| Immediately Doctor | 844 (0.3319) |
| Today Hospital | 3 (0.0012) |
| Today Doctor | 579 (0.2277) |
| Later Doctor | 36 (0.0142) |
| Today Call Centre | 26 (0.0102) |
| Later call centre | 60 (0.0236) |
| Today Pharmacy | 0 (0) |
| Later Pharmacy | 30 (0.0118) |
| Today Selfcare | 0 (0) |
| Later Selfcare | 77 (0.0303) |
| Unclear | 2 (0.0008) |

Tab. 4. Distribution of cases according to the various triage levels, as defined in Tab. 2 (N = 2543)

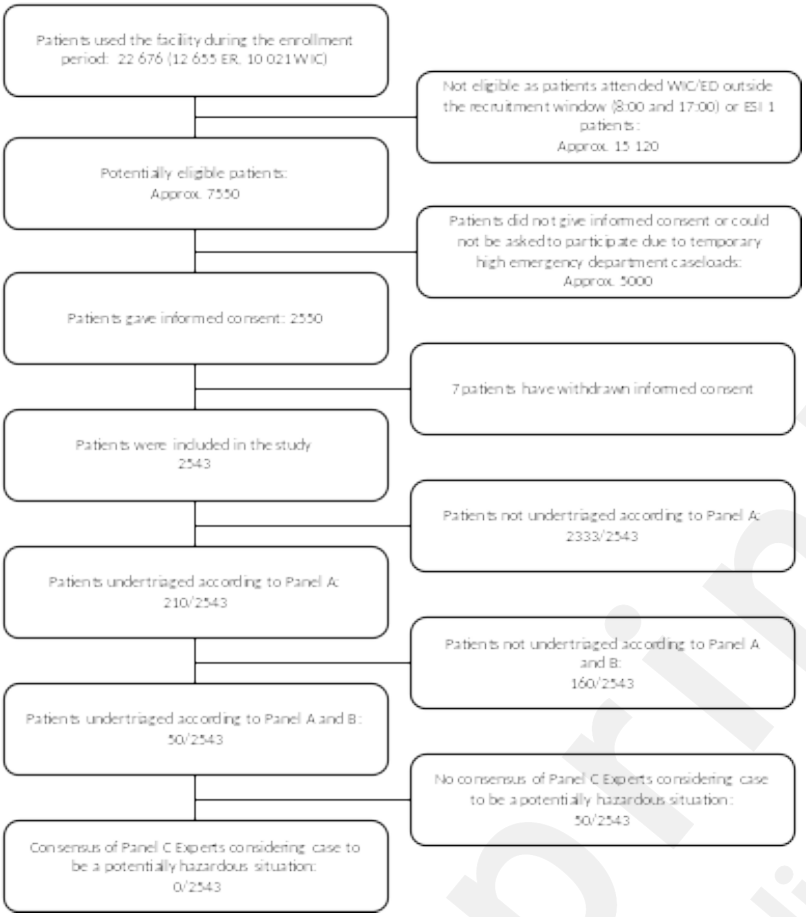


Fig 1. Flow of patients through the study and triage assessment steps by expert Panels A, B and C

In 210 (8.3%) of the 2543 cases, the recommendation issued by the symptom-checker was below the range of appropriate triage levels defined by the Panel A experts and therefore undertriaged. 50 (1.9%) of these 210 patients were equally undertriaged according to Panel B. However, in none of these 50 patients did Panel C reach a consensus that the undertriage was potentially hazardous. This resulted in an upper 95% confidence bound for the probability of a potentially hazardous undertriage of 0.1184%. If the criterion for potentially hazardous undertriage was defined as a majority of Panel C members considering life-threatening or harmful self-triage “rather likely” or “likely”, four of the 50 cases fulfilled this criterion. This resulted in an upper 95% confidence bound for the probability of a potentially hazardous undertriage of 0.3616%. Table 5 shows the adjudication of potentially hazardous undertriage for all 50 cases evaluated by the experts of Panel C.

| Case Number | Unlikely | Rather Unlikely | Rather Likely | Likely |
|---|---|---|---|---|
| Case 1 | 2 | 3 | | |

| | | | | |
|---|---|---|---|---|
| Case 2 | 5 | | | |
| Case 3 | 3 | 2 | | |
| Case 4 | 3 | 2 | | |
| Case 5 | 4 | | 1 | |
| Case 6 | 5 | | | |
| Case 7 | 2 | 2 | 1 | |
| Case 8 | 4 | 1 | | |
| Case 9 | | 2 | 2 | 1 |
| Case 10 | 4 | | 1 | |
| Case 11 | 5 | | | |
| Case 12 | 5 | | | |
| Case 13 | 1 | 1 | 2 | 1 |
| Case 14 | 4 | 1 | | |
| Case 15 | 2 | 2 | 1 | |
| Case 16 | 5 | | | |
| Case 17 | 4 | 1 | | |
| Case 18 | 5 | | | |
| Case 19 | 5 | | | |
| Case 20 | 5 | | | |
| Case 21 | 3 | 1 | 1 | |
| Case 22 | 2 | 1 | 1 | 1 |
| Case 23 | 3 | 1 | 1 | |
| Case 24 | 5 | | | |
| Case 25 | 5 | | | |
| Case 26 | 5 | | | |
| Case 27 | 5 | | | |
| Case 28 | 5 | | | |
| Case 29 | 2 | 1 | 2 | |
| Case 30 | 4 | | 1 | |
| Case 31 | 5 | | | |
| Case 32 | 2 | 1 | 1 | 1 |
| Case 33 | 4 | 1 | | |
| Case 34 | 4 | 1 | | |
| Case 35 | 5 | | | |
| Case 36 | | 2 | 3 | |
| Case 37 | 4 | 1 | | |
| Case 38 | 5 | | | |
| Case 39 | 5 | | | |
| Case 40 | 2 | 1 | 1 | 1 |
| Case 41 | 1 | 3 | 1 | |
| Case 42 | 5 | | | |
| Case 43 | 5 | | | |
| Case 44 | 5 | | | |
| Case 45 | 5 | | | |
| Case 46 | 5 | | | |
| Case 47 | 5 | | | |
| Case 48 | 1 | | 4 | |

| | | | | |
|---|---|---|---|---|
| Case 49 | 4 | 1 | | |
| Case 50 | 5 | | | |

Tab. 5. Distribution of assessment for potentially hazardous undertriage for all 50 cases, as adjudicated by each of the five members of Panel C

The central (two-sided) 95% Clopper-Pearson confidence interval for the probability of undertriage according to Panel A is [7.22%, 9.40%]. The central (two-sided) 95% Clopper-Pearson confidence interval for the probability of overtriage according to Panel A (450 cases, 17.7%) is [16.23%, 19.24%].

For the 50 out of 2543 cases that were undertriaged according to the judgments of Panels A and B, the central (two-sided) 95% Clopper-Pearson confidence interval for the corresponding probability is [1.539%, 2.688%].

The central (two-sided) 95% Clopper-Pearson confidence interval for the probability of a potentially hazardous undertriage for the consensus criterion (zero out of 2543 cases) is [0.0000%, 0.1458%] and [0.0431%, 0.4045%] according to majority criterion (four out of 2543 cases).

**Discussion**

Our study corroborates the safety of SMASS Pathfinder symptom-checker for medical self-assessment of acute complaints in a real-life clinical setting. A stepwise evaluation of 2543 consecutive patients by three independent expert panels yielded no cases of potentially hazardous undertriage when the concensus criterion was applied and four cases when the majority criterion was applied.

In a hitherto unpublished systematic literature search, we have found insufficient evidence from comparatively small studies for a safe use of symptom-checkers in clinical routine [15]. One study with 825 patients showing “exactly matched” triage in 52.6% has been published in abstract form only [16]. Another study yielded correct triage in only 50-74% [17]. A third study from Germany evaluated the safety of urgency advice provided to 378 patients at an interdisciplinary emergency department by a symptom-checker [18], showing undertriage in 34 (8.9%) and overtriage in 216 (57.1%) cases. A potentially hazardous situation was identified in 20 (5.3%) cases. This figure appears considerably higher than our finding, although an interrater variability was not taken into account in the German study. Another study aimed to analyze the performance of a clinical decision support system that allowed patients to self-triage in the emergency department of a university hospital. The authors concluded that the self-triage device was safe, as the assessments by the system and the physicians were congruent with respect to the classification as an emergency. However, in contrast to our study, the risk to life or health was not assessed [19].

In the absence of a broad study base, we cannot compare our results with previous, similarly designed studies for symptom-checkers. In contrast, medical telephone triage has been extensively evaluated during the last 25 years [20] [21] [22] [23] [24] [25] and has gained broad clinical support, despite ambivalent conclusions regarding safety.

A systematic review analysing 13 observational studies and 10 studies that simulated high-risk patients, safe triage was found in 46% to 97% [26]. Another systematic review involving computer-assisted telephone triage in urgent care [27] pointed out four studies that indicated potential undertriage errors [28] [29] [30] [31]. Notably, hospitalisation rates of patients who were advised to seek non-urgent care ranged from 9.2% to 48%. Potentially life-threatening situations emerged in 0.84% of cases according to one study [30].
We have previously investigated the safety of computer-assisted telephone triage in 208 patients with non-life-threatening conditions consulting the emergency department at a university hospital [32]. We found poor

agreement between the assessments by the call centre, the emergency physician, and the general practitioners who later cared for the patients. In one case, a risk to health or life was found. The Cochrane Collaboration in their 2004 systematic review on telephone triage concluded that insufficient data existed regarding safety [33]. In light of the available information, the results of our study compare favourably to the published data on telephone triage.

Our study has several strengths and weaknesses. We included a large number of patients in a real-world clinical setting. In addition, the study design enabled us to eliminate the low inter-rater reliability of medical triage decisions by having three independent expert panels. This allows robust conclusions about the safety of the evaluated symptom-checker.

For reasons of feasibility, we performed our study in a hospital setting, where patients were triaged to the WIC or emergency department according to ESI criteria. Thus, a wide variety of cases could be assessed. On the other hand, the symptom-checker was not utilized in a setting outside the hospital, limiting generalizability. However, presenting symptoms largely overlap with those encountered in primary care, and a potential selection bias toward more severe cases would support the conclusion on the device's safety if it were used in primary care.

A potential limitation of our study is its single-centre design. However, the Cantonal hospital Baden serves a mixed urban and rural population of ~300 000 and offers all medical services with the exception of cardiac surgery and neurosurgery. We therefore believe that the patient sample in our study is fairly representative of the general population.

The total number of patients frequenting the WIC and ED during the time of recruitment was 22 676; thus, only ~11% of them participated in the study, potentially resulting in selection bias. Due to limited resources, inclusions were possible only during the daytime, leaving ~7550 potential participants. Also, 1.5% of the 22 676 were ESI 1 patients, who were not eligible for the study. It could be speculated that patients visiting an emergency department at night time might be more seriously ill than those during the daytime. This potential bias would make our cohort more comparable to a setting in primary care.

In our study, we have focused on the safety of the symptom-checker. A possible limitation may have resulted from the fact that each case was initially assessed by a single member of Panel A. This could have precluded passing a potentially hazardous case to Panel B. While maximum patient safety may theoretically be desirable, it should be weighed against the disadvantages of overtriage, notably inefficiency, unnecessary

referrals and a higher risk of overmedicalisation, all of which increase costs. In our study, the overtriage rate after assessment by Panel A was 17.7% (450 cases). This figure is comparable to published rates of overtriage by teleconsultation and teletriage, which range from 12% to 57% [34] [35] [36] [37] [38]. In a further round of data analysis, we will also have the overtriaged cases assessed by Panel B in order to include the low interrater reliability in the analysis. As with undertriaged cases, this is likely to reduce the overtriaged cases.

From the end of February, the COVID-19 pandemic required special hygiene measures for the tablet computers used, making patient recruitment more difficult as the first wave of the pandemic peaked in March 2020. The pandemic is also likely to have affected the case mix, which may have shifted slightly towards COVID 19-positive patients.

The urgency grading used in the two-dimensional matrix for the triage levels (Table 1) was defined at the discretion of the study team, implicating a certain degree of subjectiveness. While the range of appropriate triage levels was defined based on this order, the experts did not always explicitly mark all of the intermediate triage levels as appropriate.

**Conclusion**

The SMASS pathfinder symptom-checker proved to be a safe triage tool, avoiding undertriage in a real-life clinical setting of emergency consultations at a walk-in clinic and emergency department. Although for practical reasons the symptom-checker was not evaluated outside the hospital environment, our data do not suggest that its safety may have been compromised if used for self-triage by patients in a domestic setting.

## Authors Contributions

AM, JR, MS and MV were responsible for the concept and design. PR, MS and BG did the acquisition of the data. Which were analysed and interpreted by MV, AM, JD, JR. The first draft was written by AM, and then critically revised it in conjunction with JR, JD and MV. Data analysis was done by MV, with oversight by AM and JR.

## Conflicts of interests

Andreas Meer is the founder and CEO of in4medicine. Bettina Grunder is a part-time employee of in4medicine, and Jacopo Demurtas received a scientific grant from in4medicine Inc. Michael Vock received a honorarium from in4medicine. The other authors report no conflicts of interest.

## Funding

The study was funded by the Health Innovation Hub of the Cantonal Hospital Baden, Switzerland. The symptom-checker was provided by in4medicine, Inc., at no charge.

## Acknowledgments

We thank the study nurses of the Clinical Trial Unit, Simone Fontana, Stefanie Leuenberger, and Franziska Rutz for their excellent work as well as Prof Elena Righi and Prof Lee Smith for proofreading the manuscript.

## Data Availability

The raw data of our study is available from:
https://pub.in4medicine.ch/fileadmin/pub/2024_05_01_Data_ms_58157.pdf

**References**


1. Thierrin C, Augsburger A, Dami F, Monney C, Staeger P, Clair C. Impact of a telephone triage service for non-critical emergencies in Switzerland: A cross-sectional study. PLoS One. 2021 Apr 2;16(4):e0249287. doi: 10.1371/journal.pone.0249287. PMID: 33798216.
2. Steeman L, Uijen M, Plat E, Huibers L, Smits M, Giesen P. Out-of-hours primary care in 26 European countries: an overview of organizational models. Fam Pract. 2020 Nov 28;37(6):744-750. doi: 10.1093/fampra/cmaa064. PMID: 32597962.
3 REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0745
4 Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet. 2018 Nov 24;392(10161):2263-2264. doi: 10.1016/S0140-6736(18)32819-8. Epub 2018 Nov 6. PMID: 30413281.
5. 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: 10.2196/43803. PMID: 37266983; PMCID: 10276326.
6. El-Osta A, Webber I, Alaa A, Bagkeris E, Mian S, Taghavi Azar Sharabiani M, Majeed A. What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. BMJ Open. 2022 Apr 27;12(4):e053566. doi: 10.1136/bmjopen-2021-053566. Erratum in: BMJ Open. 2022 Jul 15;12(7):e053566corr1. PMID: 35477872; PMCID: PMC9047920.
7. Munro J, Clancy M, Knowles E, Sampson F, Nicholl J. Evaluation of NHS direct: impact and appropriateness. Final report of the phase 1 research. 2001. Available from: https://www.researchgate.net/publication/246112002_Evaluation_of_NHS_Direct_first_wave_sites_Final_report_of_the_phase_1_research (accessed 02 May 2024).
8. Snooks H, Peconi J, Munro J, Cheung WY, Rance J, Williams A. An evaluation of the appropriateness of advice and healthcare contacts made following calls to NHS Direct Wales. BMC Health Serv Res. 2009 Sep 30;9:178. doi: 10.1186/1472-6963-9-178. PMID: 19793398; PMCID: PMC2761899.
9. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ. 2015 Jul 8;351:h3480. doi: 10.1136/bmj.h3480. PMID: 26157077.

10. Chambers D, Cantrell A, Johnson M, Preston L, Baxter SK, Booth A, Turner J. Digital and online symptom checkers and assessment services for urgent care to inform a new digital platform: a systematic review. Southampton (UK): NIHR Journals Library; 2019 Aug. PMID: 31433612.

11. Christ M, Bingisser R, Nickel CH. Bedeutung der Triage in der klinischen Notfallmedizin [Emergency Triage. An Overview]. Dtsch Med Wochenschr. 2016 Mar;141(5):329-35. German. doi: 10.1055/s-0041-109126. Epub 2016 Mar 3. PMID: 26939102.

12. St George I, Cullen M, Branney M. Healthline: do primary care doctors agree with the advice? N Z Med J. 2005 Oct 28;118(1224):U1693. PMID: 16258573.

13. Gribben B. General practitioners' assessments of the primary care caseload in Middlemore Hospital Emergency Department. N Z Med J. 2003 Feb 21;116(1169):U329. PMID: 12601406.

14. Durand AC, Gentile S, Gerbeaux P, Alazia M, Kiegel P, Luigi S, Lindenmeyer E, Olivier P, Hidoux MA, Sambuc R. Be careful with triage in emergency departments: interobserver agreement on 1,578 patients in France. BMC Emerg Med. 2011 Oct 31;11:19. doi: 10.1186/1471-227X-11-19. PMID: 22040017; PMCID: PMC3215166.

15. Demurtas J, Veronese N, Fittipaldo A, Smith L, Meer A. Efficacy and Safety of symptom checkers for self triage in primary care: a systematic review and meta-analysis. PROSPERO 2021 CRD42021277509 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021277509 (accessed March 7, 2024)

16. Koskela T, Liu V, Kaila M. How Does Triage by an Electronic Symptom Checker Match with Triage by a Nurse? Stud Health Technol Inform. 2022 May 25;294:571-572. doi: 10.3233/SHTI220528. PMID: 35612149.

17. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung SC, Leung LP. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. *Hong Kong J Emerg Med* 2019. DOI:10.1177/1024907919842486.

18. 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 Mar 28;10(3):e32340. doi: 10.2196/32340. PMID: 35343909.

19 Määttä J, Lindell R, Hayward N, Martikainen S, Honkanen K, Inkala M, Hirvonen P, Martikainen TJ. Diagnostic Performance, Triage Safety, and Usability of a Clinical Decision Support System Within a University Hospital Emergency Department: Algorithm Performance and Usability Study. JMIR Med Inform.

2023 Aug 31;11:e46760. doi: 10.2196/46760. PMID: 37656018; PMCID: PMC10501486.

20. Derkx HP, Rethans JJ, Muijtjens AM, Maiburg BH, Winkens R, van Rooij HG, Knottnerus JA. Quality of clinical aspects of call handling at Dutch out of hours centres: cross sectional national study. BMJ. 2008 Sep 12;337:a1264. doi: 10.1136/bmj.a1264. PMID: 18790814.

21. Derkx HP, Rethans JJ, Maiburg BH, Winkens RA, Muijtjens AM, van Rooij HG, Knottnerus JA. Quality of communication during telephone triage at Dutch out-of-hours centres. Patient Educ Couns. 2009 Feb;74(2):174-8. doi: 10.1016/j.pec.2008.08.002. Epub 2008 Oct 8. PMID: 18845413.

22. Niemann S, Meer A, Simonin C, Abel T. Medical telephone triage and patient behaviour: How do they compare? Swiss Med Wkly. 2004 Mar 6;134(9-10):126-31. doi: 10.4414/smw.2004.10276. PMID: 15106022.

23. Lattimer V, George S, Thompson F, Thomas E, Mullee M, Turnbull J, Smith H, Moore M, Bond H, Glasper A. Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial. The South Wiltshire Out of Hours Project (SWOOP) Group. BMJ. 1998 Oct 17;317(7165):1054-9. doi: 10.1136/bmj.317.7165.1054. PMID: 9774295.

24. Campbell JL, Fletcher E, Britten N, Green C, Holt TA, Lattimer V, Richards DA, Richards SH, Salisbury C, Calitri R, Bowyer V, Chaplin K, Kandiyali R, Murdoch J, Roscoe J, Varley A, Warren FC, Taylor RS. Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster-randomised controlled trial and cost-consequence analysis. Lancet. 2014 Nov 22;384(9957):1859-1868. doi: 10.1016/S0140-6736(14)61058-8. Epub 2014 Aug 3. Erratum in: Lancet. 2014 Nov 22;384(9957):1848. PMID: 25098487.

25. Murdoch J, Varley A, Fletcher E, Britten N, Price L, Calitri R, Green C, Lattimer V, Richards SH, Richards DA, Salisbury C, Taylor RS, Campbell JL. Implementing telephone triage in general practice: a process evaluation of a cluster randomised controlled trial. BMC Fam Pract. 2015 Apr 10;16:47. doi: 10.1186/s12875-015-0263-4. PMID: 25887747.

26. Huibers L, Smits M, Renaud V, Giesen P, Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. Scand J Prim Health Care. 2011 Dec;29(4):198-209. doi: 10.3109/02813432.2011.629150. PMID: 22126218.

27. Sexton V, Dale J, Bryce C, Barry J, Sellers E, Atherton H. Service use, clinical outcomes and user experience associated with urgent care services that use telephone-based digital triage: a systematic review. BMJ Open. 2022 Jan 3;12(1):e051569. doi: 10.1136/bmjopen-2021-051569. PMID: 34980613.

28. Foster J, Jessopp L, Chakraborti S. Do callers to NHS Direct follow the advice to attend an accident and

emergency department? Emerg Med J. 2003 May;20(3):285-8. doi: 10.1136/emj.20.3.285. PMID: 12748156.

29. Sprivulis P, Carey M, Rouse I. Compliance with advice and appropriateness of emergency presentation following contact with the HealthDirect telephone triage service. Emerg Med Australas. 2004 Feb;16(1):35-40. doi: 10.1111/j.1742-6723.2004.00538.x. PMID: 15239753.

30. Dale J, Higgins J, Williams S, Foster T, Snooks H, Crouch R, Hartley-Sharpe C, Glucksman E, Hooper R, George S. Computer assisted assessment and advice for "non-serious" 999 ambulance service callers: the potential impact on ambulance despatch. Emerg Med J. 2003 Mar;20(2):178-83. doi: 10.1136/emj.20.2.178. PMID: 12642540

31. Stewart B, Fairhurst R, Markland J, Marzouk O. Review of calls to NHS Direct related to attendance in the paediatric emergency department. Emerg Med J. 2006 Dec;23(12):911-4. doi: 10.1136/emj.2006.039339. PMID: 17130596.

32. Meer A, Gwerder T, Duembgen L, Zumbrunnen N, Zimmermann H. Is computer-assisted telephone triage safe? A prospective surveillance study in walk-in patients with non-life-threatening medical conditions. Emerg Med J. 2012 Feb;29(2):124-8. doi: 10.1136/emj.2009.080614. Epub 2010 Oct 20. PMID: 20961939.

33. Bunn F, Byrne G, Kendall S. Telephone consultation and triage: effects on health care use and patient satisfaction. Cochrane Database Syst Rev. 2004 Oct 18;(4):CD004180. doi: 10.1002/14651858.CD004180.pub2. PMID: 15495083.

34. Brasseur E, Gilbert A, Donneau AF, Monseur J, Ghuysen A, D'Orio V. Reliability and validity of an original nurse telephone triage tool for out-of-hours primary care calls: the SALOMON algorithm. Acta Clin Belg. 2022 Jun;77(3):640-646. doi: 10.1080/17843286.2021.1936353. Epub 2021 Jun 3. PMID: 34081571.

35. Nørøxe KB, Huibers L, Moth G, Vedsted P. Medical appropriateness of adult calls to Danish out-of-hours primary care: a questionnaire-based survey. BMC Fam Pract. 2017 Mar 14;18(1):34. doi: 10.1186/s12875-017-0617-1. PMID: 28292257.

36. Cook R, Thakore S, Morrison W, Meikle J. To ED or not to ED: NHS 24 referrals to the emergency department. Emerg Med J. 2010 Mar;27(3):213-5. doi: 10.1136/emj.2008.064261. PMID: 20304891.

37. Giesen P, Ferwerda R, Tijssen R, Mokkink H, Drijver R, van den Bosch W, Grol R. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? Qual Saf Health Care. 2007 Jun;16(3):181-4. doi: 10.1136/qshc.2006.018846. PMID: 17545343.

38. Scarfone RJ, Luberti AA, Mistry RD. Outcomes of children referred to an emergency department by an after-hours call center. Pediatr Emerg Care. 2004 Jun;20(6):367-72. doi: 10.1097/01.pec.0000133610.42699.41. PMID: 15179144.

# Supplementary Files

# Figures

Flowchart patient recruiting and triage assessment Panel A, B and C.