The Triage Capability of Laypersons: Retrospective Exploratory Analysis

Marvin Kopka[1,2], MSc; Markus A Feufel[3], Dipl-Ing (FH), MSc, PhD; Felix Balzer[1], MSc, MD, PhD; Malte L Schmieding[1], MD

[1]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Cognitive Psychology and Ergonomics, Department of Psychology and Ergonomics (IPA), Technische Universität Berlin, Berlin, Germany

[3]Division of Ergonomics, Department of Psychology and Ergonomics (IPA), Technische Universität Berlin, Berlin, Germany

Corresponding Author:
Marvin Kopka, MSc
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 581 052
Email: marvin.kopka@charite.de

Abstract

Background: Although medical decision-making may be thought of as a task involving health professionals, many decisions, including critical health–related decisions are made by laypersons alone. Specifically, as the first step to most care episodes, it is the patient who determines whether and where to seek health care (triage). Overcautious self-assessments (ie, overtriaging) may lead to overutilization of health care facilities and overcrowded emergency departments, whereas imprudent decisions (ie, undertriaging) constitute a risk to the patient's health. Recently, patient-facing decision support systems, commonly known as symptom checkers, have been developed to assist laypersons in these decisions.

Objective: The purpose of this study is to identify factors influencing laypersons' ability to self-triage and their risk averseness in self-triage decisions.

Methods: We analyzed publicly available data on 91 laypersons appraising 45 short fictitious patient descriptions (case vignettes; N=4095 appraisals). Using signal detection theory and descriptive and inferential statistics, we explored whether the type of medical decision laypersons face, their confidence in their decision, and sociodemographic factors influence their triage accuracy and the type of errors they make. We distinguished between 2 decisions: whether emergency care was required (decision 1) and whether self-care was sufficient (decision 2).

Results: The accuracy of detecting emergencies (decision 1) was higher (mean 82.2%, SD 5.9%) than that of deciding whether any type of medical care is required (decision 2, mean 75.9%, SD 5.25%; $t_{>90}$=8.4; P<.001; Cohen d=0.9). Sensitivity for decision 1 was lower (mean 67.5%, SD 16.4%) than its specificity (mean 89.6%, SD 8.6%) whereas sensitivity for decision 2 was higher (mean 90.5%, SD 8.3%) than its specificity (mean 46.7%, SD 15.95%). Female participants were more risk averse and overtriaged more often than male participants, but age and level of education showed no association with participants' risk averseness. Participants' triage accuracy was higher when they were certain about their appraisal (2114/3381, 62.5%) than when being uncertain (378/714, 52.9%). However, most errors occurred when participants were certain of their decision (1267/1603, 79%). Participants were more commonly certain of their overtriage errors (mean 80.9%, SD 23.8%) than their undertriage errors (mean 72.5%, SD 30.9%; $t_{>89}$=3.7; P<.001; d=0.39).

Conclusions: Our study suggests that laypersons are overcautious in deciding whether they require medical care at all, but they miss identifying a considerable portion of emergencies. Our results further indicate that women are more risk averse than men in both types of decisions. Layperson participants made most triage errors when they were certain of their own appraisal. Thus, they might not follow or even seek advice (eg, from symptom checkers) in most instances where advice would be useful.

XSL•FO
RenderX

Introduction

Background

Increased emergency department (ED) crowding and longer waiting times are associated with higher mortality [1,2], increased adverse events [3], and worse patient outcomes in general [4]. To mitigate these problems, each patient's urgency is assessed upon arrival in the ED—a process called triage. Usually, triage decisions are made by nurses and trained specialists, whose workload is generally high [5]. Increased workload and overcrowding pose huge risks as triage errors currently occur in roughly 16% of all cases [6] and nurses often rely on their intuition and speed up triage by not collecting further information, for example, medical history and physiological data [7,8]. These risks and the workload for health care workers could be reduced if patients were guided to appropriate health care facilities before visiting an ED to receive an initial assessment.

Getting patients to visit the health care facility most suitable for their symptoms can, for example, be achieved by offering phone triage, a telephone hotline patients can call to get a remote urgency assessment of their symptoms. For example, Roivainen et al [9] found that giving medical advice was sufficient for one-third of callers with nonemergent cases, and ED workload subsequently decreased by 36%. Another study by Midtbø et al [10] found that ED attendance decreased from 68.7% to 23.4% when advertising telephone triage hotlines. However, the accuracy of triage hotlines seems to be only around 71% with undertriaging occurring in 12% of all cases [11]—which could be potentially dangerous for some patients. Furthermore, a study based on commercial claims data from 2011 to 2013 concludes that direct-to-consumer telehealth may increase health care utilization and costs by making access to health care more convenient. Thus, such services may shift and increase, rather than reduce, the demand for health care services [12].

Another solution to disburden health care services lies in empowering patients to self-assess their medical complaints and thereby improving their ability to adequately decide which type of health care facility to visit (ie, self-triage), or where appropriate to stay at home and care for themselves. Since these decisions are made by laypersons instead of medical professionals, they come with various challenges. For example, previous studies indicate that laypersons tend to overtriage [13] and that women rate symptoms as more urgent than men [14] and are thus potentially even more inclined toward overtriage. Moreover, an Australian study showed that although laypersons are risk averse when making triage decisions, they cannot reliably detect emergencies either [15].

To address these deficits, decision support systems are designed to aid laypersons in their self-triage decision-making process, for example, kiosks in the ED [13] or symptom checkers [16]. When using them, around 25% of all patients seem to have reduced perceived urgency of their complaints, and in an experimental study, most participants also followed the advice received by the symptom checkers [17,18]. However, the accuracy of these systems varies widely [19-22]. Implementing such decision aids raises several new questions: can laypersons adequately judge when to seek such decision support? Would all laypersons profit the same way from such decision support tools? Which are the most challenging decisions to laypersons, that is, where is advice from a decision support tool most needed?

Objective

To provide a foundation for tackling these questions, in this study we investigate whether laypersons' triage accuracy and their risk averseness differ by sociodemographic variables and whether laypersons can potentially gauge whether they require advice in their decision-making, by exploring if laypersons' confidence in their triage appraisals functions as a reliable predictor of accuracy.

Methods

Ethical Considerations

No ethics approval was required for this study. Approval of the original study [23] was granted by the Ethics Committee of the Department of Psychology and Ergonomics (IPA) at Technische Universität Berlin (Tracking number: FEU_03_20180615). Participants volunteered to participate in the survey, and informed consent was obtained.

Data Collection

This analysis builds on data collected in a previous study by Schmieding et al [23], which was made available in a public open data repository [24]. They compared laypersons' triage capabilities with symptom checker performance, which are tools developed to provide clinical decision support to laypersons. Their study found that laypersons' overall triage accuracy was mediocre (mean 60.9%, SD 6.8%) based on a set of 45 fictitious patient descriptions (case vignettes). These 45 case vignettes were originally compiled by Semigran et al [21], whose study reported a mediocre overall triage accuracy for a sample of 15 symptom checkers, which was similar to laypersons' accuracy in Schmieding et al's study [23].

The layperson sample consisted of 91 US residents without prior professional medical training. They were recruited from Amazon Mechanical Turk (MTurk) in March 2020. Participants were paid US $4 for completing the web-based survey and assessing the 45 case vignettes unaided. As an incentive, a bonus of US $3 was rewarded if they achieved an accuracy above 58%. Compared with the US general population, the layperson sample had a higher level of education (all participants had at least a high school degree) and included a higher proportion of men (55/91, 60.4%) than women (36/91, 39.5%).

A more detailed description of the participants' characteristics is provided by Schmieding et al [23]. Here, we use their data set to further explore individual differences influencing laypersons' triage assessment and decisions.

Web-Based Survey

Schmieding et al [23] developed a web-based survey in which participants were asked to rate the urgency of presented case vignettes. They adapted 45 case vignettes from Semigran et al [21], comprising 15 cases for each of the 3 triage levels (self-care, nonemergency, or emergency care). The vignettes, as chosen by Semigran et al [21], included both common and uncommon chief complaints from a wide range of diagnoses and were collected from various clinical sources, including teaching materials for health care professionals.

After every triage assessment, participants were asked how certain they were in their assessment on a 4-point Likert-scale ("Very uncertain," "Rather uncertain," "Rather certain," and "Very certain"). Three sociodemographic variables were surveyed (gender, age, and level of education) and rated on a 5-tiered ordinal scale ("Non-high school graduate," "High school graduate," "Some college," "Bachelor's degree," and "Graduate degree").

Data Analysis

We conducted analyses and generated the images using base R 4.0.5 (R Core Team) [25] and the packages ggplot2 [26], RColorBrewer [27], and tidyverse packages [28].

We dichotomized triage levels to explore whether laypersons could reliably distinguish whether emergency care was required or not (decision 1), and whether self-care was sufficient or not (decision 2). Whether and where health care should be sought are the 2 common questions symptom checkers are approached with [19,29]. For decision 1, we grouped self-care cases with nonemergency cases to assess whether participants were able to correctly detect when emergency care is necessary. For decision 2, we combined emergency and nonemergency cases to the category "health care" to verify whether participants were able to correctly assess when seeing a health care professional rather than staying at home is appropriate. For each of these binary decisions, we calculated means and standard deviations for common metrics of signal detection theory (accuracy, sensitivity, specificity, and negative and positive predictive values). When determining negative and positive predictive values (NPVs and PPVs), we used the occurrence rate within the sample of case vignettes as prevalence.

To explore which factors influence laypersons' risk averseness, triage accuracy, and confidence in their own triage appraisal,

we used linear models to quantify relationships between continuous variables (age and risk averseness) and compared proportions between subgroups for ordinal and categorical variables (gender, education, and certainty).

Overtriage errors were defined as appraising the case's urgency as more urgent than necessary (eg, the participant suggested emergency care when nonemergency care was appropriate) and undertriage errors as judging it as less urgent than required. Risk averseness was defined as the proportion of overtriage errors compared to all errors, whereas a participant's triage accuracy was calculated as the proportion of correctly solved cases related to all cases.

Results

Laypersons' Triage Capability in Binary Triage Levels

On average, the participants were able to correctly classify whether a fictitious patient required emergency care or not in about 36 out of 45 cases (mean 82.2%, SD 5.88%; see decision 1 in Figure 1). The majority of participants (85/91, 93.4%) achieved an accuracy of 75% or higher. Sensitivity for detecting emergencies (mean 67.5%, SD 16.4%) was lower than the corresponding specificity (mean 89.6%, SD 8.6%), that is, the rate of assessments where cases not requiring emergency care were correctly classified as such. The PPV for detecting emergencies (mean 78.5%, SD 11.8%) was lower than the NPV (mean 85.3%, SD 6.0%).

Concerning decision 2 (Figure 1) on whether or not professional medical care is required, the overall accuracy was lower on average (mean 75.9%, SD 5.25%). Two-thirds of participants (58/91, 63.7%) achieved an accuracy of 75% or greater concerning decision 2. Here the observed sensitivity (mean 90.5%, SD 8.3%) was higher than the corresponding specificity (mean 46.7%, SD 15.9%). The PPV for this decision (mean 77.6%, SD 4.7%) was similar to the NPV (mean 76.1%, SD 16.5%).

The difference in accuracy between decision 1 and decision 2 was found to be statistically significant in a post-hoc 2-sided t test ($t_{>90}$=8.44; P<.001) with a large effect size (Cohen d=0.88). The described patterns are similar when values for accuracy, sensitivity, and specificity are broken down by gender (Figure 1). However, male participants missed a case that required medical care 1.4 times more often than female participants (false negative rates for decision 2 of 10.8%, 178/1650, for men and 7.6%, 82/180, for women), and 1.5 times more often that a case required emergency care (false negative rates for decision 1 of 37.8%, 312/825, for men and 24.4%,132/540, for women).

Figure 1. Accuracy, sensitivity, and specificity for 2 binary triage decisions by participants' gender.
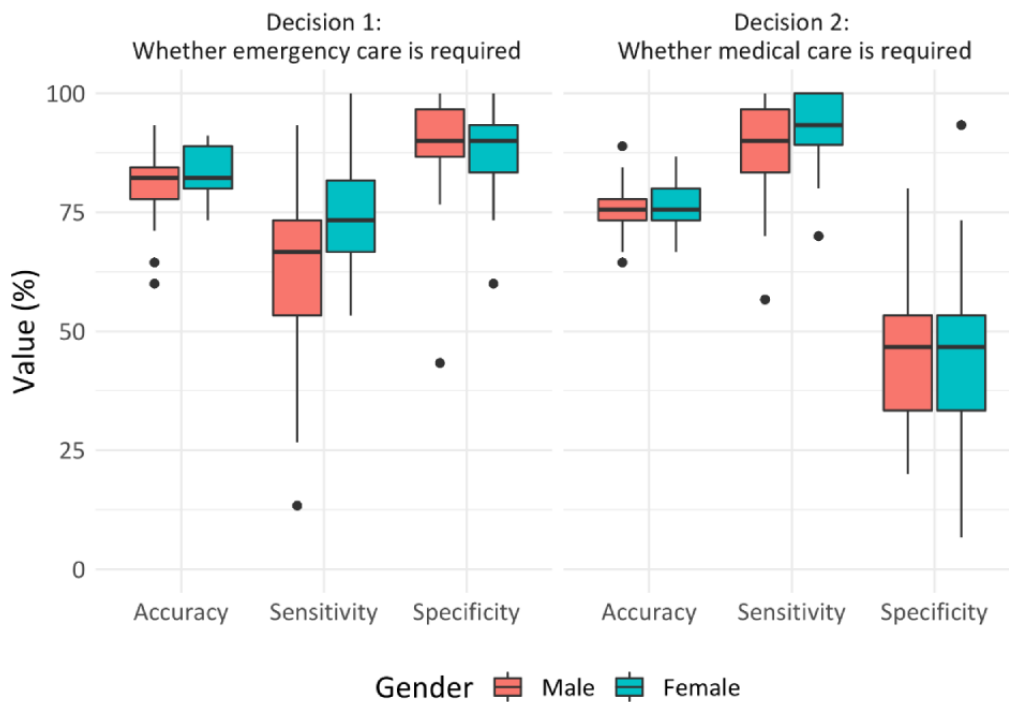
Demographic Variables and Laypersons' Risk Averseness

Participants were generally risk averse [23]. Most of their errors were overtriage errors. Only a small portion of participants (24/91, 27%) made more undertriage errors than overtriage errors.

Age only explained little variance in the decisions made (R^2=0.004; Figure 2), and median risk averseness was similar for each education level (Figure 3). Risk averseness varied with gender; that is, female participants were more risk averse than their male counterparts (Figure 4): for male participants, the ratio of overtriage to undertriage error was 1.2:1 (549:452 vignette evaluations), in contrast to 2:1 for female participants (407:195 vignette evaluations). Subsequently, the average female participant's proportion of overtriage errors among all errors was higher (mean 65.9%, SD 17.3%) than the respective proportion for male participants (mean 55.6%, SD 16.0%). This difference was found to be statistically significant in a post-hoc 2-sided Welch t test ($t_{>68.6}$=2.85; P=0.006) with a medium effect size (d=0.62).

Figure 2. Risk averseness by age.
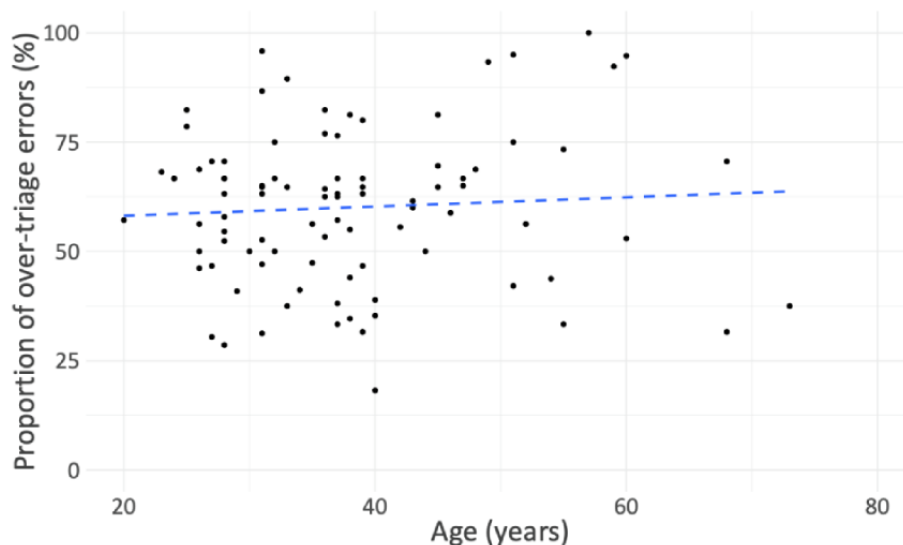
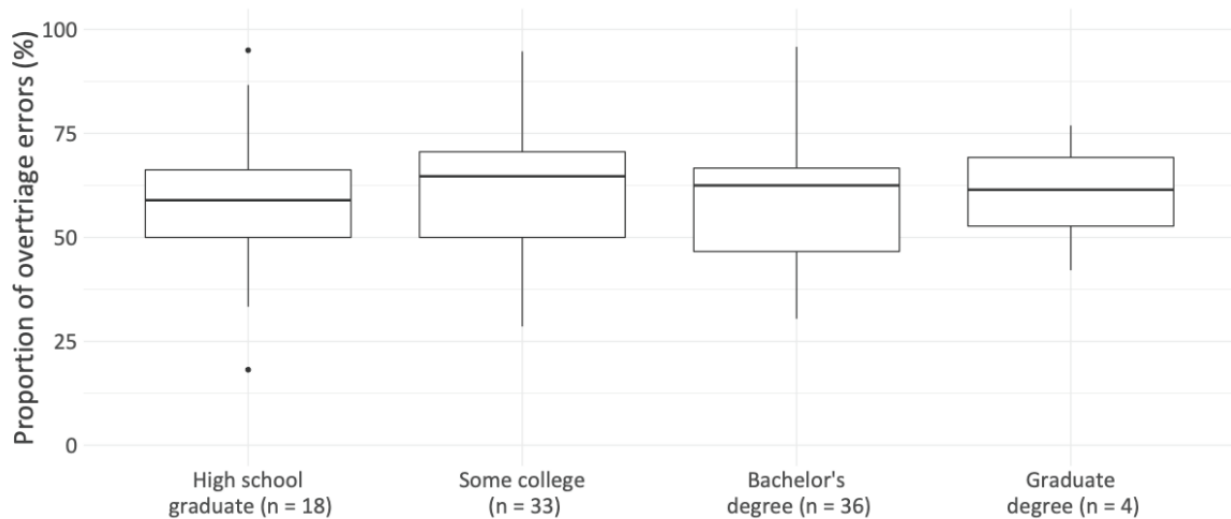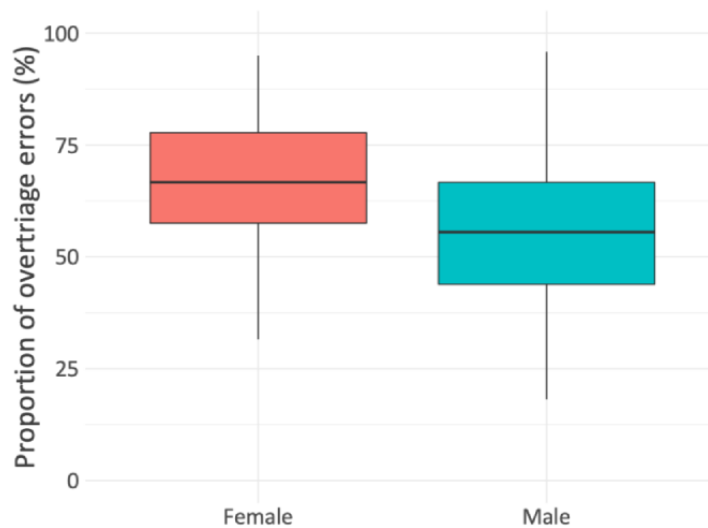**Figure 3.** Risk averseness by education.

Figure 4. Risk averseness by gender.

Decisional Certainty and Laypersons' Triage Behavior

Overall Decisional Certainty

Participants were certain in most of their triage appraisals. They indicated being "very certain" or "rather certain" in about 4 out of 5 (3381/4095, 82.6%) of the triage assessments, see Table 1 for details. Only a small portion of participants (33/91, 36.2%) indicated having been "very uncertain" in one or more of their triage decisions.

Table 1. Triage assessment and accuracy by certainty.

Triage assessment	Degree of certainty				Total
	"Very uncertain"	"Rather uncertain"	"Rather certain"	"Very certain"	
Correct	24	354	1145	969	2492
Incorrect	34	302	886	381	1603
Total	58	656	2031	1350	4095
Accuracy (%)	41.4	53.9	56.3	71.8	60.8

Decisional Certainty and Triage Accuracy

Participants' triage accuracy varied with their degree of certainty, see Table 1: it was the highest when they indicated being "very certain" (969/1350, 71.8%) and the lowest when they indicated being "very uncertain" (24/58, 41.4%). For decisions where participants were uncertain (pooling "rather uncertain" and "very uncertain"), the accuracy of their triage decision was lower (378/714, 52.9%) than when being rather or very certain (2114/3381, 62.5%; see Table 2).

When mistaken, participants were on average still commonly certain about their assessment, though less certain than when their appraisal was correct (see Tables 1 and 2). When being correct, they were certain in 84.9% (2114/2492) of the correct assessments, whereas they were certain in 79.0% (1267/1603) of all incorrect assessments. This difference is statistically significant in a post-hoc paired-sample t test ($t_{>90}$=5.43; P<.001; d=0.57).

Table 2. Triage assessment and accuracy by dichotomized certainty levels.

Degree of certainty	Triage assessment		Total	Accuracy (%)
	Correct	Incorrect		
Uncertain	378	336	714	52.9
Certain	2114	1267	3381	62.5
Total	3381	1603	4095	60.8
Proportion certain (%)	87	79	N/A[a]	N/A

[a]N/A: not applicable.

Decisional Certainty and Type of Error

Regarding the 2 types of errors, that is, overtriaging and undertriaging, we observed that participants' decisional certainty was higher when overtriaging (proportion of overtriage errors where they were either rather or very certain, 783/956, 81.9%) than when undertriaging (484/647, 74.8%). The ratio of overtriage to undertriage errors increased with the level of certainty, from 1.12:1 (18:16) for appraisals where participants were very uncertain to 2:1 (254:127) for appraisals where participants were very certain, see Table 3. The average proportion of participants being certain while overtriaging was higher (mean 80.9%, SD 23.8%) than that while undertriaging (mean 72.5%, SD 30.9%). This difference was statistically significant in a post-hoc paired-sample t test ($t_{>89}$=3.70; P<.001) with a small effect size (d=0.39).

Table 3. Triage errors by certainty.

Error type	Degree of certainty, n (%)				Total
	"Very uncertain"	"Rather uncertain"	"Rather certain"	"Very certain"	
Overtriage	18 (1.9)	155 (16.2)	529 (55.3)	254 (26.6)	956
Undertriage	16 (2.5)	147 (22.7)	357 (55.2)	127 (19.6)	647
Total	34 (2.1)	302 (18.8)	886 (55.3)	381 (23.8)	1603

Discussion

Principal Findings

As previously reported [15,23], laypersons' ability to triage is systematically biased toward overtriage, but they tend to miss emergency cases. However, when analyzing actionable metrics of triage ability, we see a majority of laypersons being fairly competent (accuracy ≥75%) in deciding whether they should seek health care (decision 2) and whether emergency care is required (decision 1). The analysis of these binary decisions also revealed a more differentiated insight into laypersons' risk averseness: they seem to be risk averse regarding decision 2 and thus tend to seek care unnecessarily, but not regarding decision 1, which suggests that they have problems identifying emergencies.

Concerning decision 1, participants were more prone to undertriage (ie, not identifying emergencies) and made only a few overtriage errors. In decision 2, they mostly made overtriage errors (ie, not recognizing that self-care is sufficient) and only a few undertriage errors. This adds further evidence to previous findings that laypersons tend to overtriage [13], but it extends them and shows that this is true only when deciding between self-care and need for a health care professional. Indeed, laypersons were more likely to undertriage when looking at emergency cases. This supports results from a study by Mills et al [15], who found that laypersons often do not recognize emergencies. It also suggests that even when assuming symptom checkers will become highly reliable and patients would use them to improve their decision-making, the benefit of these decision aids might help to disburden health care facilities of low-acuity care but not emergency care; the number of patients (rightly) presenting to the ED might increase with decision aids.

Our results indicate that women are more risk averse and overtriage more than men. This finding is in line with a previous study that found women to rate their symptoms as more urgent [14]. The two other demographic variables we examined (level of education and age) were not associated with accuracy and risk averseness.

As current symptom checkers used to assist laypersons in their triage decisions are rather risk averse [21,30], they would therefore be of greater benefit to men, who made more unsafe decisions. However, women appear to be more regular users of symptom checkers [29].

Participants' judgement of their decisional certainty predicted to some extent whether their stand-alone triage assessment was correct: when uncertain, participants were more likely to make

incorrect assessments than when being certain. At first glance, this suggests that perceived uncertainty is a good prompt for when to seek decision support. However, whether correct or incorrect, our study participants were certain about most of their judgements. Thus, perceived certainty is not a reliable predictor of triage errors, and based on their level of certainty, it seems likely that laypersons may not be able to correctly determine when they would benefit from a decision aid. Especially as participants' certainty was greater when overtriaging than when undertriaging, even perfect decision aids might not effectively reduce unnecessary doctor's visits and disburden health care facilities. This contrasts with the suggestions by Winn et al [17]. Their study found that users' perceived urgency to seek health care commonly decreased after encountering a symptom checker. Their study, however, does not consider whether the provided advice was correct. Therefore, it remains an open question, whether the use of symptom checkers (even when assuming perfect accuracy) can contribute to disburdening health care facilities.

Limitations

Limitations regarding participants and the evaluation of symptom checker accuracy are reported in detail by Schmieding et al [23] and Semigran et al [21]. Here we report the most important limitations again. Among the main limitations are that the sample of case vignettes is neither exhaustive (eg, mental health issues were excluded) nor proportionate to the incidence of diseases or medical complaints in a real-world setting. Thus, in particular, the reported values for NPVs and PPVs are not to be taken at face value, because they only reflect the prevalence in the sample of vignettes.

Second, although it has been reported that case vignettes are a valid method to assess the health care decision-making of physicians [31-33], the external validity of case vignette–based approaches with layperson decision makers has not been explored yet; that is, laypersons may decide very differently when assessing clinical vignettes compared to when they assess their own or someone else's medical complaints in the real world [21,23].

Beyond that, there are further limitations specific to the analyses in this study. Our results may have limited external validity: first, the sample was not representative of the US general population; second, triage appraisals might be influenced by the context of the health care system and by recruiting participants online; and third, population groups with no or low (information) technology affinity are not represented. Thus, future studies with more representative panels are required to determine whether our findings hold true for the broader population and whether factors other than gender and perceived certainty influence laypersons' health care decisions (eg, eHealth literacy [34], health anxiety or hypochondria [35], or propensity to trust [36]). In turn, this knowledge may help to specify for which decision and for whom decision aids would provide a benefit [34-36].

We applied statistical significance testing sparingly, since this study is a retrospective exploratory analysis and was primarily intended to generate hypotheses that can be tested (experimentally) in future studies to draw inferences.

Methodologically, we considered certainty as a measure of whether participants would consult decision aids and whether they would be open to incorporating the recommendations into their decision-making. This implies that perceived certainty in one's own appraisal correlates inversely with the openness to follow contradicting advice. However, this relationship is not based on any data. Future studies need to test this assumption.

Conclusions

Our study suggests that laypersons are overcautious in deciding whether they require medical care (decision 2). At the same time, they miss identifying a considerable proportion of emergencies (decision 1). Our results also indicate that women are more risk averse than men in both these decisions. When providing correct advice, decision aids such as symptom checkers could be of benefit to users as they could help reduce the number of missed emergencies and unnecessary visits to low-acuity care facilities. Thus, from a health system's perspective, decision aids might disburden health care facilities more of low-acuity care than of emergency care. However, layperson participants made most triage errors, and especially overtriage errors, when being certain of their own appraisal. Thus, they might not follow or even seek such advice in most instances where advice would be useful. More studies are needed to better understand laypersons' ability to self-triage, how this ability could be improved, how decision aids may support laypersons' medical decision-making, and when laypersons are willing to take advice from decision aids.

Conflicts of Interest

References

1. Lin S, Ge S, He W, Zeng M. Association of delayed time in the emergency department with the clinical outcomes for critically ill patients. QJM 2021 Aug 29;114(5):311-317. [doi: 10.1093/qjmed/hcaa192] [Medline: 32516375]
2. Woodworth L. Swamped: emergency department crowding and patient mortality. J Health Econ 2020 Mar;70:102279. [doi: 10.1016/j.jhealeco.2019.102279] [Medline: 32062054]
3. Weissman JS, Rothschild JM, Bendavid E, Sprivulis P, Cook EF, Evans RS, et al. Hospital workload and adverse events. Med Care 2007 May;45(5):448-455. [doi: 10.1097/01.mlr.0000257231.86368.09] [Medline: 17446831]

XSL•FO
RenderX

4.   Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L, Han W, et al. Effect of emergency department crowding on outcomes
      of admitted patients. Ann Emerg Med 2013 Jun;61(6):605-611.e6 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026]
      [Medline: 23218508]

5.   Innes K, Jackson D, Plummer V, Elliott D. A profile of the waiting room nurse in emergency departments: an online survey
      of Australian nurses exploring implementation and perceptions. Int Emerg Nurs 2019 Mar;43:67-73. [doi:
      10.1016/j.ienj.2018.10.003] [Medline: 30381142]

6.   Ausserhofer D, Zaboli A, Pfeifer N, Solazzo P, Magnarelli G, Marsoner T, et al. Errors in nurse-led triage: an observational
      study. Int J Nurs Stud 2021 Jan;113:103788. [doi: 10.1016/j.ijnurstu.2020.103788] [Medline: 33120136]

7.   Wolf L. Does your staff really "get" initial patient assessment? Assessing competency in triage using simulated patient
      encounters. J Emerg Nurs 2010 Jul;36(4):370-374. [doi: 10.1016/j.jen.2010.04.016] [Medline: 20624578]

8.   Wolf LA, Delao AM, Perhats C, Moon MD, Zavotsky KE. Triaging the emergency department, not the patient: United
      States emergency nurses' experience of the triage process. J Emerg Nurs 2018 May;44(3):258-266. [doi:
      10.1016/j.jen.2017.06.010] [Medline: 28750891]

9.   Roivainen P, Hoikka MJ, Raatiniemi L, Silfvast T, Ala-Kokko T, Kääriäinen M. Telephone triage performed by nurses
      reduces non-urgent ambulance missions: a prospective observational pilot study in Finland. Acta Anaesthesiol Scand 2020
      Apr 22;64(4):556-563. [doi: 10.1111/aas.13542] [Medline: 31898315]

10.   Midtbø V, Fotland SS, Johansen IH, Hunskaar S. From direct attendance to telephone triage in an emergency primary
      healthcare service: an observational study. BMJ Open 2022 May 02;12(5):e054046 [FREE Full text] [doi:
      10.1136/bmjopen-2021-054046] [Medline: 35501086]

11.   Morreel S, Colliers A, Remmen R, Verhoeven V, Philips H. How accurate is telephone triage in out-of-hours care? An
      observational trial in real patients. Acta Clin Belg 2022 Apr 30;77(2):301-306. [doi: 10.1080/17843286.2020.1839719]
      [Medline: 33124524]

12.   Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does
      not decrease spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline:
      28264950]

13.   Trivedi S, Littmann J, Stempien J, Kapur P, Bryce R, Betz M. A comparison between computer-assisted self-triage by
      patients and triage performed by nurses in the emergency department. Cureus 2021 Mar 19;13(3):e14002 [FREE Full text]
      [doi: 10.7759/cureus.14002] [Medline: 33884243]

14.   Cooper AA, Humphreys KR. The uncertainty is killing me: self-triage decision making and information availability. E-JAP
      2008 Oct 14;4(1):1-6. [doi: 10.7790/ejap.v4i1.124]

15.   Mills B, Hill M, Buck J, Walter E, Howard K, Raisinger A, et al. What constitutes an emergency ambulance call? Australas
      J Paramed 2019 Mar 22;16 This is an online article and does not have a page number. Please delete the filler page number
      0 in the XML. [doi: 10.33151/ajp.16.626]

16.   Coyle N, Kennedy A, Schull MJ, Kiss A, Hefferon D, Sinclair P, et al. The use of a self-check-in kiosk for early patient
      identification and queuing in the emergency department. CJEM 2019 Nov 06;21(6):789-792. [doi: 10.1017/cem.2019.349]
      [Medline: 31057137]

17.   Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for
      seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561]
      [Medline: 31880791]

18.   Kopka M, Schmieding ML, Rieger T, Roesler E, Balzer F, Feufel MA. Determinants of laypersons' trust in medical decision
      aids: randomized controlled trial. JMIR Hum Factors 2022 May 03;9(2):e35219 [FREE Full text] [doi: 10.2196/35219]
      [Medline: 35503248]

19.   Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker
      apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810]
      [Medline: 35536633]

20.   Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential
      impact on service utilisation. PLoS One 2021;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088]
      [Medline: 34265845]

21.   Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit
      study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]

22.   Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps
      in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]

23.   Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers
      against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475 [FREE Full text] [doi:
      10.2196/24475] [Medline: 33688845]

24.   Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Data set supplementing "Benchmarking triage
      capability of symptom checkers against that of medical laypersons: survey study". Zenodo. 2021. URL: https://zenodo.org/
      record/4454538#.YyCACXZBzIU [accessed 2022-09-13]

25.   R Core Team. R: A Language and Environment for Statistical Computing. The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. URL: https://www.R-project.org/ [accessed 2022-09-12]

26.   Wickham H. ggplot2.3.5. Vienna, Austria: R Foundation for Statistical Computing URL: https://ggplot2.tidyverse.org/ [accessed 2022-09-12]

27.   Neuwirth E. RColorBrewer: ColorBrewer Palettes. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/web/packages/RColorBrewer/index.html [accessed 2022-09-12]

28.   Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. JOSS 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]

29.   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 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]

30.   Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May 07;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]

31.   Meyer AND, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. JAMA Intern Med 2013 Nov 25;173(21):1952-1958. [doi: 10.1001/jamainternmed.2013.10081] [Medline: 23979070]

32.   Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 2000 Apr 05;283(13):1715-1722. [doi: 10.1001/jama.283.13.1715] [Medline: 10755498]

33.   Yager J, Linn LS, Leake B, Gastaldo G, Palkowski C. Initial clinical judgments by internists, family physicians, and psychiatrists in response to patient vignettes: I. Assessment of problems and diagnostic possibilities. Gen Hosp Psychiatry 1986 May;8(3):145-151. [doi: 10.1016/0163-8343(86)90072-1] [Medline: 3710146]

34.   Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]

35.   Loos A. Cyberchondria: too much information for the health anxious patient? J Consum Health Internet 2013 Oct;17(4):439-445. [doi: 10.1080/15398285.2013.833452]

36.   Jessup S, Schneider T, Alarcon G, Ryan T, Capiola A. The measurement of the propensity to trust automation. In: Chen JYC, Fragomeni G, editors. Virtual, Augmentation and Mixed Reality Applications and Case Studies. Cham: Springer International Publishing; 2019:476-489.

Abbreviations

ED: emergency department
NPV: negative predictive value
PPV: positive predictive value

XSL•FO

RenderX