The Journal of High School Research

Do AI Symptom Checkers Help or Hinder Timely Care-Seeking for Early Heart Disease? A Narrative Review of the Evidence by Insurance Status

Darren Kapturski*

Submitted: 26 September 2026 **Accepted:** 09 October 2026 **Publication date:** 10 November 2026

DOI: 10.70671/bxv8wq54

Abstract: Cardiovascular disease killed 919,032 people in the United States in 2023, and how long people wait before getting help shapes much of that toll. The wait is not the same for everyone. In a 24-hospital U.S. registry, 48.6% of uninsured heart attack patients waited more than six hours, compared with 44.6% of insured patients who worried about costs and 39.3% of insured patients who did not. AI symptom checkers now show up at the exact moment when people decide whether to get care. Yet, the author found no study that measures whether using one changes how fast people seek care for heart symptoms, or whether the answer depends on insurance status. This narrative review, organized with guidance from PRISMA 2020, brings together evidence on four topics: how accurate symptom checkers are, how they change people's plans for care, who uses them, and how insurance shapes the decision to seek cardiac care. The author searched PubMed/MEDLINE and Google Scholar between June and September 2026, along with publisher, journal, and government websites, and checked the reference lists of key studies. Twenty-two sources form the core evidence base: 20 peer-reviewed studies and 2 survey reports that were not peer-reviewed. Because the studies ranged from simulated-case audits to national registries, the author appraised each one with four questions (real patients or simulated cases, observed behavior or stated intention, independence from the tool's developer, and how representative the sample was) and synthesized the findings narratively. Following the Commonwealth Fund, the author treated insurance as three levels: adequately insured, underinsured, and uninsured. The evidence cuts both ways. Symptom checkers reduce uncertainty and usually catch emergencies. However, in the largest real-world dataset, users lowered the urgency of their care plans eight times as often as they raised it. Most tools triage no better than an average layperson, and in studies with real patients, 2–9% of cases were under-triaged. In a randomized trial, people using AI chatbots made no better care decisions than people using other sources. The typical user is insured, college-educated, and female, so little is known about the groups most likely to delay care. The author argues that insurance likely decides whether a symptom checker speeds care up or slows it down. Insured users can act on advice, or double-check it, at little cost. Uninsured and underinsured users are the least able to catch a falsely reassuring answer. Three hypotheses follow. Among adequately insured adults, prior symptom checker use will shorten the time to professional consultation for early cardiac symptoms. Among uninsured adults, it will not shorten that time and may lengthen it. Underinsured adults will fall in between, and pairing advice with navigation to low-cost care should narrow the gaps among the three groups. The author outlines an equity-focused design with measurable safety thresholds, such as sensitivity for emergencies of 95% or higher and harmful under-triage in 2% or fewer encounters, along with community health worker review and community governance. The author also proposes four studies to test the hypotheses: an audit of standardized cardiac cases, a prospective cohort study stratified by insurance tier, a stepped-wedge trial of community health worker navigation, and qualitative interviews.

Author keywords: artificial intelligence; symptom checkers; cardiovascular disease; care-seeking delay; health insurance

Introduction

Heart disease and the problem of delay

Cardiovascular disease (CVD) kills more people than anything else in the world. The World Health Organization estimates that 19.8 million people died of it in 2022, about 32% of all deaths, and more than three-quarters of those deaths happened in low- and middle-income countries.[1] In the United States, CVD killed 919,032 people in 2023, roughly one in three deaths. Care and medication for heart disease alone cost more than $168 billion in 2021 and 2022.[2]

Timing drives much of this burden. Treatment that reopens a blocked artery during a heart attack works best when it starts fast. For the most severe type, ST-elevation myocardial infarction, U.S. guidelines aim for 90 minutes or less from first medical contact to opening the artery.[3]

*Corresponding Author: Darren Kapturski.
Email: dmkapturski@gmail.com
Bergen County Technical High School, Teterboro, NJ 07608, United States; Mentor: Amol Joshi, PhD, MS, MBA, Wake Forest School of Medicine/Pioneer Academics

However, the clock starts well before anyone reaches the hospital. The Atherosclerosis Risk in Communities (ARIC) study reviewed the records of 18,928 heart attack patients. In 2000, 49.5% of them took four hours or longer to reach the hospital after their symptoms began, and Black patients and women consistently waited longer than White patients and men.[4] This delay has hardly budged in decades. In the Worcester Heart Attack Study, the median delay was about two hours in 1986 and still about two hours in 2005, and patients who waited were less likely to receive reperfusion treatment within 90 minutes of arrival.[5] A systematic review of 42 studies found the same pattern repeatedly: women and older adults reach the hospital later than men and younger adults.[6]

Heart failure has its own version of the problem. Breathlessness and ankle swelling creep up over days, not minutes, so they are easy to wait out. Among 258 heart failure patients in China, the median gap between worsening symptoms and seeking care was 7.5 days. Longer waits went along with milder breathlessness, worse swelling, depression, weak support from the health system, and caregivers who struggled to manage symptoms.[7] In an Australian study, patients who felt they had little control over their illness, or who did not see themselves as heart failure patients, tended to cope passively and wait.[8]

Across these studies, the costliest delay comes before a patient ever contacts the health system. It happens in a private moment: someone notices a symptom and has to decide whether it is worth acting on. That decision is where most preventable delay begins, so anything that speeds it up or slows it down matters clinically. AI symptom checkers were designed for exactly that moment. Whether they help people act sooner is the question this review sets out to answer.

Insurance status as a structural driver of delay

Whether someone acts on a worrying symptom often comes down to whether they can afford to. The TRIUMPH registry followed 3,721 heart attack patients at 24 U.S. hospitals. Delays longer than six hours occurred in 48.6% of uninsured patients, 44.6% of insured patients who had financial worries about getting care, and 39.3% of insured patients without those worries.[9] The result is a gradient: uninsured patients waited the longest, underinsured patients came next, and adequately insured patients waited the least. An ARIC analysis found something similar: Medicaid recipients and people in low-income neighborhoods were more likely to wait a long time, even after accounting for clinical factors and distance to the hospital.[10]

For this reason, this review treats insurance at three levels rather than a simple yes or no. The author follows the Commonwealth Fund's definitions. Adults are *adequately insured* if they had coverage all year and their out-of-pocket costs stayed affordable. They are *underinsured* if they had coverage all year, but their out-of-pocket costs or deductibles were high compared with their income. They are *uninsured* if they had no coverage.[11] In the Commonwealth Fund's 2024 national survey of working-age adults, 56% were adequately insured, 23% were underinsured, 12% had a gap in coverage during the past year, and 9% were uninsured. Among the underinsured, 57% reported that they had skipped needed care because of the cost.[11]

Where people live adds another layer. Among 2.18 million Medicare patients, those treated at rural hospitals were less likely to undergo cardiac catheterization after a heart attack (49.7% vs. 63.6%). They were also more likely to die within 30 days of a heart attack or heart failure admission than patients treated at urban hospitals.[12] Table 1 summarizes these disparities and notes what each number actually measures. Delay, access, and death rates are different outcomes, so they should not be compared directly.

Symptom checkers at the point of decision

AI symptom checkers are apps and websites that ask about a user's symptoms, suggest what might be going on, and recommend a next step: call emergency services, get seen within a day, book a routine visit, or manage things at home. Ada and Buoy Health are two widely used examples. They are free, they work at any hour, and they need no appointment. On paper, that makes them a good fit for the moment when delay begins. Recent commentaries in *JAMA* have pointed to the promise of AI in diagnosis while warning against the hype.[13,14] For symptom checkers, the harder question is about behavior, not technology: does the advice change what people do, and how quickly?

Research gap and aims

Researchers have tested how accurate symptom checkers are. Separately, other researchers have shown how insurance shapes cardiac care-seeking. Yet, almost no one has looked at where the two meet. The author found no study that measured whether using a symptom checker changes the time between the start of cardiac symptoms and a visit to a professional, and no study that compared insured, underinsured, and uninsured users. The care-seeking studies that do exist ask people what they plan to do rather than timing what they actually did; they are not specific to heart symptoms, and they do not break results down by insurance status.

This review tries to fill that gap. The main question is as follows: among adults, does using an AI symptom checker help or hinder how quickly people seek professional care for early signs of heart disease, and does insurance change the answer? The author breaks this question into four smaller questions: (1) How accurate and safe is symptom checker triage? (2) How does using a symptom checker change people's plans for care? (3) Who uses symptom checkers? (4) How does insurance affect whether someone can act on a recommendation? The author then uses these findings to sketch an equity-focused design, a governance model, and a research agenda.

Table 1. Evidence on cardiac care-seeking delay and access by population group

Group	Key finding	What it measures	Source
Women and older adults	Reached the hospital later after heart attack symptoms than men and younger adults in most of 42 studies	Prehospital delay	Nguyen et al.[6]
Black patients and women	Consistently longer delays than White patients and men; 49.5% of all patients delayed 4 hours or more (2000)	Prehospital delay	McGinn et al.[4]
Uninsured	48.6% delayed more than 6 hours after heart attack symptom onset, vs. 39.3% of the adequately insured	Prehospital delay	Smolderen et al.[9]
Underinsured	44.6% of insured patients with financial concerns delayed more than 6 hours; 57% of underinsured adults avoided needed care due to cost	Prehospital delay; access	Smolderen et al.[9], Collins and Gupta[11]
Medicaid recipients; low-income neighborhoods	Higher odds of long vs. short delay (low neighborhood income OR 1.46)	Prehospital delay	Foraker et al.[10]
Rural residents (age 65+)	Lower catheterization rates (49.7% vs. 63.6%); higher 30-day mortality (HR 1.10 for heart attack; 1.15 for heart failure)	Access; mortality	Loccoh et al.[12]
People with heart failure	Median of 7.5 days from symptom worsening to seeking care	Care-seeking delay	Xu et al.[7]
All adults	About 1 in 5 heart attacks is silent: damage occurs but is not recognized	Symptom recognition	Centers for Disease Control and Prevention [CDC][2]

Methods

Review design

This is a narrative review that borrows elements of the PRISMA 2020 reporting guidance, such as stated search terms, eligibility criteria, and study appraisal.[15] It is not a systematic review. Searches were iterative rather than registered in advance, the number of records screened at each stage was not recorded, and study selection was not checked by a second reviewer. No meta-analysis was attempted because the included studies used different designs, populations, and outcomes. These choices are discussed as limitations in the Limitations section.

Information sources and search strategy

Searches were conducted between June and September 2026 in PubMed/MEDLINE and Google Scholar, supplemented by publisher, journal, and government websites. Two core searches were used, shown here in PubMed syntax and adapted for other sources. Search A (symptom checkers): ("symptom checker" *OR "symptom assessment app*" OR "self-triage" OR "online triage" OR "digital triage") AND ("care seeking" OR "help seeking" OR "triage accuracy" OR "health care utilization" OR "intention"). Search B (cardiac delay and insurance): ("myocardial infarction" OR "heart failure" OR "chest pain" OR "cardiovascular") AND ("prehospital delay" OR "care-seeking delay" OR "patient delay") AND ("insurance" OR "uninsured" OR "underinsured" OR "Medicaid" OR "socioeconomic").

Reference lists of key studies were examined to find further studies (backward citation searching); for example, the reference list of Chan et al.[16] led to Semigran et al.[17] and Winn et al.[18] Government and professional-society sources, such as the CDC, WHO, and the Commonwealth Fund, were used for descriptive statistics and regulatory context. Every source cited was located at its original publisher, journal, or government website, and its key figures were checked against the published abstract or full text.

Eligibility criteria

Studies were included if they (a) were peer-reviewed original research, systematic reviews, or official government or professional-society reports; (b) were published in English; (c) involved adults; and (d) reported at least one of the following: the triage accuracy or safety of patient-facing symptom checkers; care-seeking intentions or behavior after symptom checker use; the characteristics of symptom checker users; or care-seeking delay for cardiac conditions by insurance status, income, or demographic group.

The author excluded studies of pediatric populations only, clinician-facing decision support tools, commentaries without original data (except when cited for context), conference abstracts without full reports, and duplicate reports. Grey literature, such as industry surveys, was included only when

no peer-reviewed source addressed the same question, and it is labeled as grey literature wherever it is cited.

Study selection

Sources were included when they met the eligibility criteria and addressed at least one of the four sub-questions. Twenty-two sources form the core evidence base for the four themes in the Results section: 20 peer-reviewed studies and 2 survey reports that were not peer-reviewed.[11,19] A further 24 sources, including clinical guidelines, design frameworks, regulatory documents, and national statistics, provide context in the Introduction and Discussion sections. Because screening counts were not recorded, a PRISMA flow diagram is not reported.

Data extraction and quality appraisal

For each included study, the author extracted the design, setting, sample size, population, tool studied, outcome, main finding, and key limitations. Because the included studies ranged from simulated-case audits to national registries, the author used four appraisal questions developed for this review rather than a single scored tool: (1) Did the study use real patients rather than simulated cases? (2) Was the outcome observed behavior or stated intention? (3) Were the authors independent of the tool developer? (4) Was the sample likely to represent the population of interest? Each core study was then rated as providing higher, moderate, or lower confidence evidence for the review question (Table 2).

Synthesis

Findings were grouped into four themes matching the sub-questions and synthesized narratively, with attention to agreements, contradictions, and gaps across studies.

Results

Overview of the included evidence

Twenty-two sources form the core evidence base (Study Selection). Table 2 summarizes the eleven core studies on symptom checkers and AI health tools. The evidence came mainly from the United States, the United Kingdom, Germany, and Canada. Most symptom checker studies used simulated cases or measured stated intentions, and none measured the timing of care-seeking for cardiac symptoms.

Theme 1: accuracy and safety of triage

Symptom checkers are only moderately accurate, and the tools vary a lot. The first large audit presented 23 checkers with the same 45 standardized cases. They provided appropriate triage advice 57% of the time: 80% for emergencies but only 33% for cases that could be handled at home, and individual tools ranged from 33% to 78%.[17] Six years later, most tools still triaged no better than an average layperson, although the best five outperformed most of the lay participants. Checkers directed self-care cases to the emergency department 24.7% of the time, compared with 4.1% for laypeople.[20] A separate UK evaluation also found wide

Table 2. Characteristics and appraisal of core studies on symptom checkers

Study	Design and setting	Key finding	Main limitation	Confidence
Semigran et al.[17]	Audit of 23 checkers using 45 standardized cases (U.S.)	Appropriate triage in 57% overall; 80% for emergencies and 33% for self-care; tools generally risk-averse	Simulated cases, not real patients	Moderate
Schmieding et al.[20]	Checkers vs. laypersons on the same cases (Germany)	Most checkers no better than laypersons; self-care cases labeled emergencies in 24.7% vs. 4.1%	Simulated cases	Moderate
Ceney et al.[21]	Evaluation of online checkers against clinical cases (UK)	Wide variation between tools; overall performance below accepted clinical standards	Simulated cases	Moderate
Chan et al.[16]	Prospective cohort, 581 patients at 2 EDs and 13 clinics (Canada)	73% accurate vs. 58% for patients; all 10 emergencies detected; potential harm from delay in 2–3%	Rule-based prototype, not machine learning; few emergencies	Higher
Cotte et al.[22]	Prospective study, 378 ED walk-in patients (Germany)	8.9% under-triaged by the Ada app; 5.3% judged potentially hazardous	Single center; lower-acuity patients only; developer co-authors	Moderate
Bean et al.[23]	Preregistered randomized study, 1,298 UK adults, 10 scenarios, 3 AI chatbots	Chatbots alone found relevant conditions in 94.9%; users with chatbots below 34.5% (conditions) and 44.2% (course of action), no better than controls	Simulated scenarios; not cardiac-specific	Higher

Table 2. (Continued)

Study	Design and setting	Key finding	Main limitation	Confidence
Winn et al.[18]	158,083 Buoy encounters with intended care before and after (U.S.)	Uncertainty fell from 34% to 21%; urgency lowered in 32%, unchanged in 65%, raised in 4%	Intentions, not behavior; appropriateness of changes unknown	Moderate
Miller et al.[24]	523 patients using Ada in a UK primary care waiting room	12.8% would have chosen lower-intensity care; 22% among ages 18–24 years	Stated intentions; authors employed by developer	Lower
Arellano Carmona et al.[25]	Survey of 2,437 Buoy users (U.S.)	88.9% insured; 81.2% college-educated; 72.5% White; 84.9% female	Self-selected respondents	Moderate
Meyer et al.[26]	Survey of 329 Isabel users (U.S.)	75.7% women; 89.1% White; many used the tool to decide whether or where to seek care	Small, self-selected sample	Lower
AXA Health[19] **(grey literature)**	Commissioned online survey of 2,000 UK adults	59% of AI users delayed care after reassurance vs. 23% of non-users; 25% received misleading advice	Not peer-reviewed; self-report; commercial sponsor	Lower

Note: ED = emergency department. Confidence reflects the four appraisal questions in the Data Extraction and Quality Appraisal section.

gaps between tools, with overall performance below what is expected elsewhere in medicine.[21]

Studies with made-up cases suggest that checkers err on the side of caution. Studies with real patients show that they still miss some urgent cases. In a Canadian cohort of 581 patients, a prototype checker outperformed patients' own judgment (73% vs. 58% accurate), caught all 10 emergencies, and would have cut hospital visits by 55%. Yet, following its advice would have risked harm from delayed care in 2–3% of encounters.[16] The authors point out that this rate is similar to that of telephone triage. In a German emergency department, the Ada app under-triaged 8.9% of 378 walk-in patients, and an expert panel judged 5.3% of cases potentially dangerous.[22] Neither study focused on heart symptoms, and both included few true emergencies. Thus, it is still unknown how often these tools under-triage early heart disease.

General-purpose AI chatbots raise the same concerns, only sharper. In a preregistered randomized study, 1,298 UK adults worked through ten medical scenarios with help from an AI chatbot (GPT-4o, Llama 3, or Command R+) or from any source they liked. On their own, the chatbots named the relevant conditions in 94.9% of cases and selected the right course of action in 56.3%. However, when ordinary people used those same chatbots, they named the relevant conditions in fewer than 34.5% of cases and selected the right course of action in fewer than 44.2%, no better than the control group.[23] The chatbots knew about medicine. The breakdown happened in the conversation between a person and a machine. For cardiac care, this means that a tool's accuracy on paper says little about whether it will help people get care faster.

Theme 2: effects on care-seeking intentions

The largest real-world dataset comes from 158,083 encounters with Buoy Health. Before and after each session, users reported what level of care they planned to seek. The share who were unsure dropped from 34% to 21%, which is a real benefit. However, among users who selected a level of care both times, 32% lowered the urgency of their plan, 65% kept it the same, and only 4% raised it.[18] In a UK primary care waiting room, 12.8% of 494 patients said the Ada app would have steered them toward less intensive care, such as self-care, a pharmacy visit, or putting off their appointment. Among 18- to 24-year-olds, this figure was 22%.[24]

At first glance, this clashes with the finding that checkers tend to over-triage made-up cases. One explanation is that people who turn to a checker often start out more worried than they need to be, so stepping down may often be the right call. Another is that real users describe their symptoms less precisely than a textbook case does, which can shift the advice downward. The existing studies cannot tell these apart, because none of them tracked what happened to patients after their plans changed. That missing link is exactly where the evidence runs out for cardiac care.

An industry survey points in the same way. In a commissioned survey of 2,000 UK adults, 59% of those who used AI to check symptoms reported that they put off contacting a health professional because the tool reassured them. Only 23% of non-users delayed after seeking other online reassurance, and 25% of AI users said they had received misleading information.[19] These numbers are self-reported and were not peer-reviewed. They are best treated as a warning sign worth studying properly, not as a measure of how big the effect really is.

Theme 3: who uses symptom checkers

The people using symptom checkers are not the most likely to delay cardiac care. Among 2,437 U.S. Buoy users, 88.9% were insured, 76.5% had a regular doctor or clinic, 81.2% had at least a college degree, 72.5% were White, and 84.9%

were female.[25] Users of the Isabel checker were 75.7% women and 89.1% White,[26] and 78% of the Buoy sessions that Winn et al.[18] analyzed involved women. Age matters too. In the UK waiting room study, 93% of 18- to 24-year-olds found the advice helpful, compared with 59% of adults aged 70 and older,[24] even though older adults face the highest cardiac risk. A scoping review of digital health tools found that they often widen existing gaps in care because access, digital skills, and trust are unevenly spread.[27]

This skews what is known. Nearly everything known about how symptom checkers affect people comes from insured, educated users who already have a regular source of care. They are exactly the people best placed to catch a bad recommendation. What the same tools do for uninsured, underinsured, rural, and older users has barely been measured.

Theme 4: insurance status and the ability to act on advice

A symptom checker's advice is only as useful as the user's ability to follow it. The TRIUMPH and ARIC findings show that insurance and income shape delay even during a heart attack, the most urgent cardiac emergency there is.[9,10] With early, vague symptoms, people have more room to weigh cost against benefit, so money probably matters even more. National survey data back this up. Among underinsured adults, 57% skipped needed care because of cost, and 41% of adults who put off care for cost reasons said a health problem got worse as a result.[11]

Research outside the symptom checker field suggests that help from a trusted person can soften these barriers. In a randomized trial of Medicaid-insured and uninsured adults with chronic conditions, a community health worker (CHW) program returned $2.47 to Medicaid for every dollar spent within a year, mostly by keeping people out of the hospital.[28] Older work points in the same direction. Low-income patients tend to seek care where they already have trusted relationships,[29] and prevention programs in low-resource settings work best when they are built into existing community structures.[30]

Answering the research question

So far, no study answers the question directly because none has timed cardiac care-seeking after symptom checker use. The evidence does support a conditional answer. Symptom checkers help by cutting uncertainty, and they usually catch emergencies. They can hurt by talking people down from urgent plans far more often than by talking them into urgent care, and by under-triaging a small but important share of cases. Whether the overall effect is faster or slower care probably depends on insurance because insurance decides what happens after the advice is given (Fig. 1).

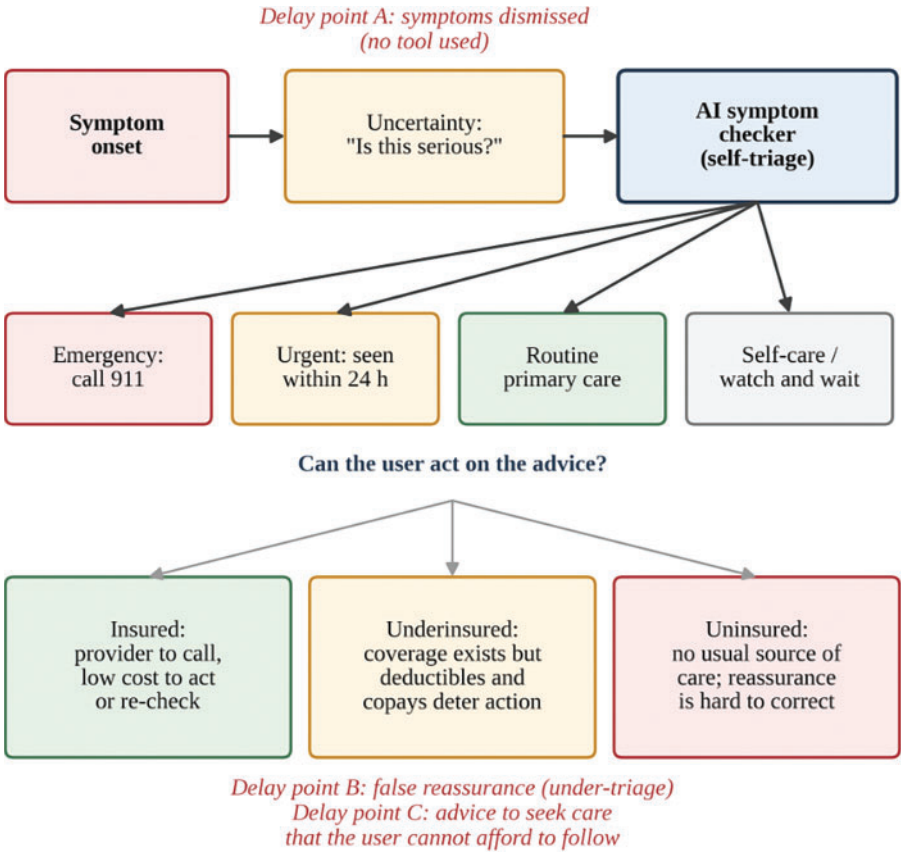


Figure 1. Where AI symptom checkers intervene in cardiac care-seeking and where delay can still occur. Note: Delay point A occurs without a tool; points B and C occur after a recommendation. The bottom row reflects the insurance gradient in Smolderen et al.[9] and Collins and Gupta[11]

Table 3. Summary of findings and certainty of evidence

Question	What the evidence shows	Main sources	Certainty	Why
1. How accurate and safe is triage?	Accuracy is modest and varies widely between tools; simulated self-care cases are over-triaged, yet 2–9% of real patients are under-triaged	Semigran et al.[17], Schmieding et al.[20], Chan et al.[16], Cotte et al.[22]	Moderate	Consistent across studies, but mostly simulated cases and none specific to cardiac symptoms
2. How does use change care-seeking?	Uncertainty falls, but intended urgency is lowered far more often than raised; people using AI chatbots decide no better than controls	Winn et al.[18], Miller et al.[24], Bean et al.[23]	Low	Mostly stated intentions; one randomized study used simulated scenarios; no study measured timing
3. Who uses symptom checkers?	Mostly insured, college-educated, White, and female users; older adults find them less helpful	Arellano Carmona et al.[25], Meyer et al.[26], Miller et al.[24]	Moderate	Consistent across three surveys, but respondents chose to take part
4. How does insurance shape acting on advice?	Uninsured and underinsured patients delay heart attack care more; cost deters care-seeking	Smolderen et al.[9], Foraker et al.[10], Collins and Gupta[11]	Moderate to high for heart attacks; low for early symptoms	Large multicenter registries, but data from 1993 to 2008 and focused on acute events
Main question: Does prior use speed or slow cardiac care, by insurance tier?	No direct evidence; the effect is likely to depend on insurance status (H1–H3)	No direct studies	Very low	No study has measured it

Note: Certainty ratings are judgments made for this review, based on study design, consistency, and how directly each study addresses the question. They are not formal GRADE ratings.

Think about what happens next. An insured user who is told to see a doctor can usually do so. An insured user who receives a falsely reassuring answer can still call their regular doctor if the symptoms do not go away. An uninsured user who is told to see a doctor runs into a cost that the app cannot remove. An uninsured user who receives false reassurance has little reason, and few affordable options, to seek a second opinion. The underinsured fall in between, which matches the delay gradient that Smolderen et al.[9] found. This leads to three hypotheses that future studies can test:

H1. Among adequately insured adults, prior symptom checker use will shorten the time to professional consultation for early cardiac symptoms, mainly by reducing uncertainty.

H2. Among uninsured adults, prior symptom checker use will not shorten, and may lengthen, the time to consultation because false reassurance is harder to correct and advice to seek care is harder to act on.

H3. Underinsured adults will show an intermediate effect, and pairing symptom checker advice with navigation to low-cost care will narrow the differences among the three groups.

Table 3 summarizes what the evidence shows for each sub-question and how much confidence it deserves.

Discussion

What existing approaches do well and where they fail

Several established approaches try to catch heart disease early. When they fall short, the problem is usually not the medicine but the logistics. Routine screening works well clinically. The SPRINT trial showed that lowering systolic blood pressure to 120 mmHg, rather than 140 mmHg, reduced major cardiovascular events.[31] But screening only reaches people who see a clinician regularly. Public education has raised awareness of heart attack symptoms, but awareness has not meant speed. In the REACT trial, an 18-month campaign in 10 U.S. cities, compared with campaigns in 10 matched cities, did not shorten the median delay of about 140 minutes, although ambulance use rose by 20%.[32] Evaluation after symptoms start works well once a patient shows up, but by design, it waits for the patient to decide to come.

Consumer devices provide hard data. In the Apple Heart Study of 419,297 participants, 0.52% received an irregular-rhythm alert from their smartwatch.[33] In the REHEARSE-AF trial, twice-weekly handheld ECG recordings using the AliveCor device found more cases of new atrial fibrillation than routine care.[34] Still, these devices mostly

pick up rhythm problems; they cost money, and the people who use them tend to be health-conscious already. The author left out newer, FDA-cleared devices for heart failure monitoring and portable 12-lead ECGs because there is not much peer-reviewed evidence on them yet. Table 4 compares these approaches on what works clinically, where each breaks down in practice, and whether it touches the pre-diagnosis delay gap: the stretch between noticing a symptom and first contacting care.

A proposed equity-focused design: cardiocheck

To make these recommendations concrete, the author sketches a hypothetical tool called CardioCheck. It does not exist and has not been tested; it is included only to make the review's recommendations specific enough to test. Its job is not to diagnose anyone, but to shorten the time between the first symptom and the right care, especially for people with the fewest other options.

Two contrasting scenarios

Picture two people who report the same symptoms: two weeks of unusual tiredness, swollen ankles, and getting winded on the stairs. Both receive the same advice, to be seen within 24 hours, because the symptoms could point to heart failure.

The first person is a 52-year-old woman in the suburbs with insurance through her job. She calls her doctor's office the next morning, gets in within two days, and blood tests confirm early heart failure. For her, the tool mostly removes doubt. Instead of waiting weeks to see if things get better, she gets care in a few days.

The second person is a 58-year-old uninsured man who lives 40 miles from the nearest hospital and has no regular doctor. A generic checker gives him the same advice, but he cannot easily follow it, since a clinic visit and tests could cost several hundred dollars. Now suppose the tool also points him to a federally qualified health center (FQHC) that charges on a sliding scale, and a community health worker calls to help him book a visit and find a ride. Suddenly, the same advice becomes something he can act on. Without that help, the advice may just confirm his fear of the bill, so he waits. The algorithm is identical in both cases. What differs is everything around it.

Evidence, proposal, and decision rights

Not every part of this design stands on the same footing. Some functions rest on current evidence: symptom checkers reduce uncertainty,[18] and some tools are highly sensitive to emergencies.[16] Others are proposals that would need validation first, including cardiac-specific questions that cover atypical symptoms, conservative thresholds, navigation to low-cost care, and CHW follow-ups. Human-centered AI design aims for tools that combine high automation with high human control, so that the result is reliable, safe, and trustworthy.[35,36] In practice, that means setting decision rights in advance. Patients decide whether to use the tool and what to share. Emergency advice appears immediately, with a one-tap button to call 911. A CHW reviews moderate-risk results within 4 hours. Uncertain answers default upward

Table 4. Existing approaches: clinical value, operational failure points, and relevance to the pre-diagnosis delay gap

Approach	What works clinically	Where it fails operationally	Addresses the pre-diagnosis delay gap?	Source
Public education campaigns	Improves symptom awareness	Did not shorten delay in a community trial; reaches already health-engaged people	Partially: raises awareness but does not support the individual decision	Luepker et al.[32]
Routine screening	Finds and treats risk factors; fewer cardiovascular events	Requires a regular clinician; cost deters uninsured and underinsured adults	No: acts before symptoms, not at symptom onset	SPRINT Research Group[31]
Symptom-triggered clinical evaluation	Rapid treatment is highly effective once patients arrive	Passive: depends on the patient deciding to come	No: begins only after the decision to seek care	O'Gara et al.[3]
Wearables and handheld ECG devices	Detects atrial fibrillation; objective data	Cost; focused on heart rhythm; users skew engaged and insured	Partially: only for rhythm symptoms and device owners	Perez et al.[33], Halcox et al.[34]
Telephone nurse triage	Established safety benchmark: harm from under-triage in 1.3–3.2% of calls	Staffing and hours vary; people must know the service exists	Yes, where available	As summarized in Chan et al.[16]
AI symptom checkers (current)	Reduce uncertainty; generally sensitive to emergencies	Variable accuracy; users lower urgency; users skew insured	Yes in principle, but the effect on timing is unmeasured	Winn et al.[18], Schmieding et al.[20]

Table 5. Proposed performance thresholds and their benchmarks

Requirement	Proposed threshold	Benchmark from the literature
Sensitivity for emergencies (cardiac cases, simulated and real)	95% or higher	80% across 23 checkers[17]; 10 of 10 emergencies[16]
Under-triage with potential for harm	2% or fewer encounters	2–3%[16]; 5.3%[22]; telephone triage 1.3–3.2% (as summarized in Chan et al.[16])
Self-care cases wrongly sent to emergency care	15% or fewer	24.7% for checkers vs. 4.1% for laypersons[20]
Performance gap across subgroups (sex, age, race and ethnicity, insurance, and language)	Emergency sensitivity within 5 percentage points of the overall rate	Clinical AI datasets overrepresent a few high-income countries[37]; a widely used algorithm underestimated Black patients' needs[38]
Performance with real users, not only the tool alone	Users reach the correct course of action within 10 percentage points of the tool's own rate	Tool alone 56.3% vs. users below 44.2%[23]
Direction of plan changes for cardiac symptoms	No subgroup with a net downward shift that is not supported by outcomes	32% lowered urgency vs. 4% raised[18]
Validation before launch	Prospective study with real patients, including uninsured and rural users	Most current evidence is simulated or intention-based (Table 2)
Monitoring after launch	Quarterly safety reports; public annual audit	No comparable requirement for consumer checkers today

to seeking care, and diagnosis and treatment stay with clinicians, who also set the rules and write the emergency protocol in advance. With the user's consent, a symptom summary can be sent to the receiving clinic.

Reliable, safe, and trustworthy: measurable requirements

This reliable, safe, and trustworthy (RST) framing is useful only if it is tied to numbers a tool must meet before and after launch. The proposed thresholds (Table 5) are drawn from the benchmarks reported in the included studies. They are targets for validation, not demonstrated properties.

Costs, savings, and sustainability

Any free tool must be paid for by someone. The author found no peer-reviewed estimate of what it costs to build and validate a symptom checker, so development ($3–8 million) and yearly maintenance ($1–2 million) are treated as illustrative assumptions that a pilot would need to replace with real data. One CHW team costs about $500,000 per year, and CHW support returned $2.47 per $1 invested within one year (range $1.84–$3.09).[28] Heart failure hospitalization costs a median of $15,879 per patient per year,[39] so avoiding these costs for about 126 patients per year would offset $2 million in annual operating costs.

The break-even calculation shows what a pilot would need to demonstrate, not what it would achieve. Funding could combine professional–society sponsorship, federal innovation and small-business research grants, and integration into FQHC services; however, these sources are proposals that require confirmation. Fig. 2 summarizes the design: the AI is the deck of a bridge between symptom uncertainty and trusted, affordable care, but human oversight is what holds the bridge up.

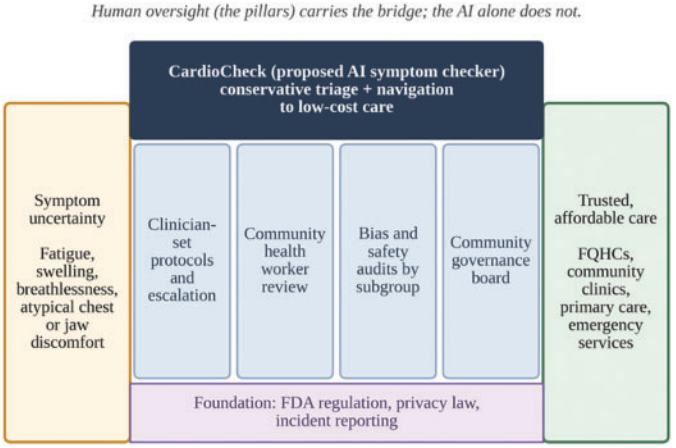


Figure 2. CardioCheck as a bridge between symptom uncertainty and trusted care, supported by human oversight. Note: Conceptual model informed by Shneiderman[35,36]. The design is proposed and has not been validated.

Risks, adoption barriers, and mapped safeguards

No design removes risk entirely. Each major risk the author identified, including practical barriers to adoption, is paired

in Table 6 with the safeguard intended to reduce it, and the author ranked the risks by their likely severity for patients.

The first three risks matter most because each one directly lengthens the delay for the people this review cares about most. Under-triage comes first because its harm cannot be undone, and because the only studies with real patients show it happens in a small but real share of cases. Uneven accuracy comes second, since errors bunched up in one group would widen the very gaps in Table 1. Advice that people cannot act on comes third because, for uninsured users, even a correct recommendation may not shorten the wait. All three risks point to the same lesson. A cautious algorithm helps, but it is not enough; the human support that follows a recommendation is what turns advice into care.

Governance and accountability

In the United States, software that gives patients specific triage recommendations may be regulated by the Food and Drug Administration (FDA) as a medical device, and the FDA has also described additional oversight for AI and machine-learning software that changes after deployment.[42,43] Many consumer health apps fall outside HIPAA but are covered by the Federal Trade Commission's Health Breach Notification Rule.[44] In Europe, the Artificial Intelligence Act places many health-related AI systems in its high-risk category.[45] Commentators have argued that responsible use of AI in medicine requires transparency, evaluation in real-world practice, and clear accountability.[46]

Regulation sets the floor. Governance decides how a tool behaves day to day. Patients, caregivers, and community organizations should take part in governance, not merely use the tool. The proposed board has 11 voting seats: 3 patients or caregivers (at least one uninsured or underinsured), 2 community organization representatives, 2 community health workers, 2 clinicians, 1 data scientist, and 1 privacy and ethics specialist. Patient, caregiver, community organization, and CHW seats hold a majority (7 of 11), and any change to safety thresholds requires a vote that includes at least one patient member. The board would review flagged cases monthly, receive subgroup performance reports quarterly, and publish an independent external audit each year. Missing any Table 5 threshold for two consecutive quarters, or any death or serious harm linked to the tool's advice, would trigger a tiered action: fix and retest within 30 days, suspend the affected symptom pathway, or pause

Table 6. Principal risks ranked by severity, with the safeguard mapped to each

Rank	Risk	Why it matters	Safeguard mapped to this risk
1	Under-triage and false reassurance	Delay in a time-critical condition can be fatal; under-triage occurred in 2–9% of real-patient encounters[16,22]	Conservative defaults; 95% emergency-sensitivity threshold; CHW follow-up for moderate risk; one-tap emergency call
2	Uneven accuracy across groups	Clinical AI datasets are unrepresentative[37]; a widely used algorithm underestimated Black patients' needs[38]	Diverse training and validation data; results reported by subgroup; bias audits before and after launch[40]
3	Advice that cannot be acted on	Underinsured and uninsured adults avoid care because of cost[11]	Navigation to sliding-scale FQHCs; CHW help with booking and transport
4	Privacy breaches	About 275 million U.S. health records were breached in 2024[41]	Collect only necessary data; encryption; no sale of data; independent security audits
5	Limited digital access and skills	Older and lower-income adults use and benefit less[24,27]	Plain language; multiple languages; a phone-based option
6	Health anxiety and repeated checking	37% of AI users reported feeling more anxious after checking symptoms (AXA Health[19]; grey literature)	One clear action at the end of each check; prompt to talk to a person after repeated checks
7	Burden on clinic workflows	Extra referrals and messages can overload busy clinics	Route moderate-risk cases to CHWs rather than physicians; share short, structured summaries only
8	Unclear liability	Unclear responsibility discourages adoption and weakens accountability	Clear statement of limitations; incident reporting; responsibilities defined by the governance board (see Governance Section)
9	Low public trust	Users may ignore advice from an unfamiliar tool	Launch through trusted community institutions[29,30]; community governance

Table 7. Proposed research agenda

Design	Population and comparison groups	Primary outcomes	Equity-specific hypothesis	Purpose
Cardiac case audit	40 or more standardized cardiac cases varying by sex, age, and typical vs. atypical symptoms, run through current checkers and AI chatbots	Emergency sensitivity; under-triage rate by the case's sex and age	Under-triage will be higher for women and older adults with atypical symptoms	Fast, low-cost baseline
Prospective cohort in EDs and FQHCs	Adults with possible cardiac symptoms; prior checker users vs. non-users, stratified into adequately insured, underinsured, and uninsured	Time from symptom onset to first professional contact; appropriateness of care setting; 30-day major adverse cardiac events	H1–H3: benefit for the insured, no benefit or harm for the uninsured, and intermediate effect for the underinsured	Tests the core research question
Stepped-wedge cluster randomized trial	FQHCs and rural clinics adopting a checker with vs. without CHW navigation	Time to consultation; visits completed within 7 days; cost per hospitalization avoided	Navigation will narrow the insured–uninsured gap in time to consultation	Tests the proposed design
Qualitative interviews	Uninsured, underinsured, rural, older, and non-English-speaking adults	Reasons for acting or not acting on advice; trust	Cost and trust, not accuracy, will be the main barriers to acting on advice	Explains the mechanism

the tool and notify regulators and users. Continued funding would depend on audit compliance, every session would state the tool's limits, and anonymized findings from each harm review would be published.

Limitations

The evidence has five main weaknesses. First, none of the studies the author included timed care-seeking after symptom checker use, and none focused on heart symptoms; therefore, this review's conclusions about cardiac care are inferred from general triage studies. Second, most of the care-seeking evidence comes from what people say they will do, which may not match what they actually do. Third, several studies were conducted or co-authored by the companies that developed the tools.[22,24] Fourth, the users in these studies were mostly insured, educated, White women, so the findings may not apply to the groups this review worries about most. Fifth, some key delay data are old (from 1987 to 2008), before current insurance policies and today's AI tools, and the AXA findings are from grey literature.

Symptom checkers are also updated frequently, so accuracy estimates from 2015 to 2022 may not describe current versions, and the only randomized study of AI chatbots used simulated scenarios that were not cardiac-specific.[23]

This review has its own limitations. It is a narrative review rather than a registered systematic review: the searches were iterative, screening counts were not recorded, study selection was not checked by a second reviewer, only English-language sources were included, and no meta-analysis was possible. Relevant studies may therefore have been missed. Finally, CardioCheck is a proposal: its thresholds, costs, and governance model are untested, and its development and maintenance costs are illustrative assumptions.

Recommendations for future work

The hypotheses in the section "Answering the Research Question" can be tested directly, and Table 7 outlines four complementary studies that together would do it.

Across all of these studies, results should be reported separately by insurance tier, sex, age, race and ethnicity, rurality, and language. The same control variables recur throughout the cardiac-delay literature: age, sex, race and ethnicity, insurance, income, education, rurality, comorbidities such as diabetes, and having a regular source of care. Each study should measure them the same way so that results can be compared.

Conclusion

Does prior use of an AI symptom checker help or hinder the speed of seeking care for early heart disease? Nobody has measured it directly yet. The evidence that does exist supports a conditional answer. Symptom checkers reduce uncertainty and usually catch emergencies. However, they lower users' intended urgency far more often than they raise it. They also occasionally miss urgent cases, and in the only randomized test, they did not help people make better care decisions.

Insurance status is likely to decide which of these effects wins out. Uninsured and underinsured patients already wait longer during heart attacks than the adequately insured.[9] They are the least able to catch false reassurance, and they

are the least represented among today's symptom checker users.[25] A tool that helps insured users might do little for the people who need faster care most. It might even hurt them.

Symptom checkers meant to reduce cardiac delay should have to meet measurable safety standards, be tested on real patients across subgroups, come paired with navigation and community health worker support, and be governed with patients and communities at the table. Above all, someone needs to measure how long people take to reach care after using a symptom checker, broken down by insurance status. Then, the next review of this question can answer it with data instead of inference.

Declarations

Acknowledgements

The author thanks Professor Amol Joshi of Pioneer Academics for mentorship and detailed feedback, as well as peer reviewers Wangshu (Hermia) Zhao, Jiayi Wan, Jinwoo Park, Jack Signorelli, and Jay Yu for their comments on earlier drafts.

Conflicts of interest

The author declares no conflicts of interest.

Funding

This research received no external funding.

Data availability

No new data were generated. All sources are listed in the references.

Use of AI tools

Zotero (Corporation for Digital Scholarship) was used to organize and format references.

References

[1] World Health Organization. Cardiovascular diseases (CVDs) [fact sheet]; 2025. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] Centers for Disease Control and Prevention. Heart disease facts; 2025. https://www.cdc.gov/heart-disease/data-research/facts-stats/index.html.

[3] O'Gara PT, Kushner FG, Ascheim DD, et al. ACCF/AHA guideline for the management of ST-elevation myocardial infarction. *Circulation*. 2013;127(4):e362–e425.

[4] McGinn AP, Rosamond WD, Goff DC Jr, Taylor HA, Miles JS, Chambless L. Trends in prehospital delay time and use of emergency medical services for acute myocardial infarction: experience in 4 US communities from 1987–2000. *Am Heart J*. 2005;150(3):392–400. doi:10.1016/j.ahj.2005.03.064.

[5] Saczynski JS, Yarzebski J, Lessard D, et al. Trends in prehospital delay in patients with acute myocardial infarction (from the Worcester Heart Attack Study). *Am J Cardiol*. 2008;102(12):1589–1594. doi:10.1016/j.amjcard.2008.07.056.

[6] Nguyen HL, Saczynski JS, Gore JM, Goldberg RJ. Age and sex differences in duration of prehospital delay in patients with acute myocardial infarction: a systematic review. *Circ: Cardiovasc Qual Outcomes*. 2010;3(1):82–92. doi:10.1161/circoutcomes.109.884361.

[7] Xu M, Ruan T, Huang X, et al. Care-seeking delay of patients with heart failure in China: a mixed-method study. *ESC Heart Fail*. 2024;11(4):2086–2099. doi:10.1002/ehf2.14757.

[8] Ivynian SE, Ferguson C, Newton PJ, DiGiacomo M. The role of illness perceptions in delayed care-seeking in heart failure: a mixed-methods study. *Int J Nurs Stud*. 2024;150(3):104644. doi:10.1016/j.ijnurstu.2023.104644.

[9] Smolderen KG, Spertus JA, Nallamothu BK, et al. Health care insurance, financial concerns in accessing care, and delays to hospital presentation in acute myocardial infarction. *JAMA*. 2010;303(14):1392–1400. doi:10.1001/jama.2010.409.

[10] Foraker RE, Rose KM, McGinn AP, et al. Neighborhood income, health insurance, and prehospital delay for myocardial infarction: the atherosclerosis risk in communities study. *Arch Intern Med*. 2008;168(17):1874–1879. doi:10.1001/archinte.168.17.1874.

[11] Collins SR, Gupta A. The state of health insurance coverage in the U.S.: Findings from the Commonwealth Fund 2024 Biennial Health Insurance Survey. Commonwealth Fund; 2024. doi:10.26099/byce-qc28.

[12] Loccoh EC, Joynt Maddox KE, Wang Y, Kazi DS, Yeh RW, Wadhera RK. Rural-urban disparities in outcomes of myocardial infarction, heart failure, and stroke in the United States. *J Am Coll Cardiol*. 2022;79(3):267–279. doi:10.1016/j.jacc.2021.10.045.

[13] Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. *JAMA*. 2023;330(4):317–318. doi:10.1001/jama.2023.11440.

[14] Wachter RM, Brynjolfsson E. Will generative artificial intelligence deliver on its promise in health care? *JAMA*. 2024;331(1):65–69. doi:10.1001/jama.2023.25054.

[15] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.

[16] Chan F, Lai S, Pieterman M, et al. Performance of a new symptom checker in patient triage: Canadian cohort study. *PLoS One*. 2021;16(12):e0260696. doi:10.1371/journal.pone.0260696.

[17] Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480. doi:10.1136/bmj.h3480.

[18] 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;2(12):e1918561. doi:10.1001/jamanetworkopen.2019.18561.

[19] AXA Health. AXA Health research shows AI is driving Britons to delay care, and seek it unnecessarily [press release; survey conducted by Censuswide]; April 14, 2026. https://www.axahealth.co.uk/news/2026/axa-health-research-shows-ai-is-driving-people-to-delay-care/.

[20] 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;23(3):e24475. doi:10.2196/24475.

[21] Ceney A, Tolond S, Glowinski A, et al. Accuracy of online symptom checkers and the potential impact on service utilisation. *PLoS One*. 2021;16(7):e0254088. doi:10.1371/journal.pone.0254088.
[22] Cotte F, Mueller T, Gilbert S, et al. 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 and uHealth*. 2022;10(3):e32340. doi:10.2196/32340.
[23] Bean AM, Payne RE, Parsons G, et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. *Nat Med*. 2026;32(2):609–615. doi:10.1038/s41591-025-04074-y.
[24] Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. *JMIR Hum Factors*. 2020;7(3):e19713. doi:10.2196/19713.
[25] Arellano Carmona K, Chittamuru D, Kravitz RL, et al. Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. *J Med Internet Res*. 2022;24(8):e36322. doi:10.2196/36322.
[26] 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. doi:10.2196/14679.
[27] Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. *J Med Internet Res*. 2022;24(3):e34144. doi:10.2196/34144.
[28] Kangovi S, Mitra N, Grande D, Long JA, Asch DA. Evidence-based community health worker program addresses unmet social needs and generates positive return on investment. *Health Aff*. 2020;39(2):207–213. doi:10.1377/hlthaff.2019.00981.
[29] Klaiman TA, Valdmanis VG, Bernet P, Moises J. Opportunities for prevention: assessing where low-income patients seek care for preventable coronary artery disease. *Popul Health Manag*. 2015;18(5):337–341. doi:10.1089/pop.2014.0122.
[30] Nishtar S. Cardiovascular disease prevention in low resource settings: lessons from the Heartfile experience in Pakistan. *Ethn Dis*. 2003;13(2 Suppl 2):S138–S148.
[31] SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015;373(22):2103–2116. doi:10.1056/NEJMoa1511939.
[32] Luepker RV, Raczynski JM, Osganian S, et al. Effect of a community intervention on patient delay and emergency medical service use in acute coronary heart disease: the Rapid Early Action for Coronary Treatment (REACT) trial. *JAMA*. 2000;284(1):60–67. doi:10.1001/jama.284.1.60.
[33] Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med*. 2019;381(20):1909–1917. doi:10.1056/nejmoa1901183.
[34] Halcox JPJ, Wareham K, Cardew A, et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. *Circulation*. 2017;136(19):1784–1794. doi:10.1161/CIRCULATIONAHA.117.030583.
[35] Shneiderman B. Human-centered artificial intelligence: reliable, safe & trustworthy. *Int J Hum-Comput Interact*. 2020;36(6):495–504. doi:10.1080/10447318.2020.1741118.
[36] Shneiderman B. *Human-Centered AI*. Oxford, UK: Oxford University Press; 2022.
[37] Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities: a global review. *PLOS Digit Health*. 2022;1(3):e0000022. doi:10.1371/journal.pdig.0000022.
[38] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. doi:10.1126/science.aax2342.
[39] Urbich M, Globe G, Pantiri K, et al. A systematic review of medical costs associated with heart failure in the USA (2014–2020). *PharmacoEconomics*. 2020;38(11):1219–1236. doi:10.1007/s40273-020-00952-0.
[40] Nazer LH, Zatarah R, Waldrip S, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health*. 2023;2(6):e0000278. doi:10.1371/journal.pdig.0000278.
[41] HIPAA Journal. 2024 healthcare data breach report; 2025. https://www.hipaajournal.com/2024-healthcare-data-breach-report/.
[42] U.S. Food and Drug Administration. Artificial intelligence/-machine learning (AI/ML)-based software as a medical device (SaMD) action plan; 2021. https://www.fda.gov/media/145022/download.
[43] U.S. Food and Drug Administration. Clinical decision support software: guidance for industry and Food and Drug Administration staff; 2022. https://www.fda.gov/media/109618/download.
[44] Federal Trade Commission. Health Breach Notification Rule (16 C.F.R. Part 318); 2024. https://www.ftc.gov/legal-library/browse/rules/health-breach-notification-rule.
[45] European Parliament, Council of the European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). OJ L, 2024/1689. 12 July 2024:1–144. http://data.europa.eu/eli/reg/2024/1689/oj.
[46] Dorr DA, Adams L, Embi P. Harnessing the promise of artificial intelligence responsibly. *JAMA*. 2023;329(16):1347–1348. doi:10.1001/jama.2023.2771.