JAMIA Open, 2024, **7(4)**, ooae102
https://doi.org/10.1093/jamiaopen/ooae102
Research and Applications

Research and Applications

Concordance of data collected by an app for medical history taking and in-person interviews from patients in primary care

Carla Joos[1], Klara Albrink[1], Eva Hummers, MD[1], Frank Müller, MD[1,2], Kai Antweiler, PhD[3], Dominik Schröder, PhD[1,*], Eva Maria Noack, PhD[1]

[1]Department of General Practice, University Medical Center Göttingen, 37073 Göttingen, Germany, [2]Department of Family Medicine, College of Human Medicine, Michigan State University, Grand Rapids, MI 49503, United States, [3]Department of Medical Statistics, University Medical Center Göttingen, 37073 Göttingen, Germany

*Corresponding author: Dominik Schröder, MSc, Department of General Practice, University Medical Center Göttingen, Humboldtallee 38, 37073 Göttingen, Germany (dominik.schroeder@med.uni-goettingen.de)

D. Schröder and E.M. Noack shared senior authorship.

Abstract

Objective: This study investigates the concordance of patient information collected using a medical history app compared to in-person interviews.

Materials and Methods: In this cross-sectional study we used an app to collect medical data from patients in family practice in Germany. Collected information included age, height, weight, perceived severity of complaints, and 38 current complaints. Subsequently, in-person interviews based on the query structure of the app were conducted with patients directly after the patient finished filling out the app. Concordance was assessed as exact matches between the data collected app-based and in-person interviews, with the in-person interview as a reference. Regression analysis examined which patient characteristics were associated with mismatching and underreporting of complaints.

Results: Three hundred ninety-nine patients were included in the study. Concordance of reported age, weight, and height, as well as perceived severity of complaints ranged from 76.2% to 96.7%. Across all 38 complaints, 64.4% of participants showed completely identical complaint selection in app-based and in-person interviews; 18.5% of all participants overreported; and 17.0% underreported at least 1 complaint when using the app. Male sex, higher age, and higher number of stated complaints were associated with higher odds of underreporting at least one complaint in the app.

Discussion: App-collected data regarding age, weight, height, and perceived severity of complaints showed high concordance. The discordance shown concerning various complaints should be examined regarding their potential for medical errors.

Conclusion: The introduction of apps for gathering information on complaints can improve the efficiency and quality of care but must first be improved.

Trial registration: The study was registered at the German Clinical Trials Register No. DRKS00026659 registered November 3, 2021. World Health Organization Trial Registration Data Set, https://trialsearch.who.int/Trial2.aspx?TrialID=DRKS00026659

Lay Summary

We developed an app to help doctors record patient information before the consultation. We tested the app to see if the information collected with the app was identical to the information collected in a face-to-face interview. To do this, we asked 399 patients to use the app and then had a face-to-face interview with study staff asking the same questions. We found that nearly all patients reported their age, weight, height, and severity of their symptoms identically in the app. However, when reporting current complaints, there was a discrepancy between the app and the face-to-face interview: 18.5% of patients reported additional complaints (overreporting); 17% reported fewer complaints in the app than in the face-to-face interview (underreporting); 64.4% gave completely identical answers for all 38 complaints. This shows that medical history apps can be a useful tool for capturing patient information, but that they need to be carefully assessed and improved before they can be used in daily practice. We recommend further research to understand the reasons for the discrepancy between the app and the face-to-face interview and to determine the medical relevance of these results.

Key words: medical history taking; digitization; application software; clinical decision-support systems; mHealth.

Introduction

Medical history taking is considered "the most powerful, sensitive and versatile tool at the doctor's disposal."[1] An accurate, comprehensive, and symptom-oriented medical history allows one to identify many medical conditions without further examinations or diagnostic tests. However, time constraints are jeopardizing adequate medical history taking. In Germany, the average doctor-patient encounter in family practice lasts only 9 min,[2] including physical examinations, diagnostic procedures, and decision-making additionally to

Received: April 2, 2024; **Revised:** August 26, 2024; **Editorial Decision:** September 19, 2024; **Accepted:** September 23, 2024

medical history taking. To address this issue, digital tools can be used to help family practitioners collect patient information, reducing errors and alleviating time constraints while also improving patient experience.[2–5] Interruptions or the use of medical language may confuse patients leading to incomplete descriptions of their complaints.[6] Tools that avoid these issues can enhance patient comfort and improve provider-patient interaction. Furthermore, a digital medical history taking may be less prone to provider bias[7] allowing for a more objective and accurate understanding of the patient's condition.

Previous studies have examined the concordance of self-reported medical data from different specialties and using different methods. Self-reported data were collected using either paper questionnaires or Internet surveys.[8–10] Data from in-person interviews or medical records were used for comparison. The medical data collected by apps have not been assessed for concordance with other data sources; instead, apps were only checked for usability and efficiency of care.[11–15] However, adequate concordance of collected data is a prerequisite for the adoption of new tools for medical history taking.

In this study, we examine the concordance of data collected using a medical history-taking app designed for use in family practice compared to in-person interviews. We examined data on age, weight, height, perceived severity of complaints, and selected acute complaints.

Methods

This study is part of the project "Digitally assisted system to obtain patients' medical history before consultation" (DASI); additional information on the methodology can be found in the study protocol.[16]

Software application

We used an app designed to take a symptom-oriented medical history directly from patients in family practice. Patients enter their sex, age, height, and weight, perceived severity of complaints, select one or more complaints, and are then guided through a dynamic questionnaire.

The app uses plain language and a straightforward user interface. After completing the query, the app compiles a short report that can be transferred into the patients' electronic health records.

For this study, the app ran on an iPad Mini 4 (Apple Inc, Cupertino, CA, USA).

Study design and recruitment

We carried out a cross-sectional study recruiting patients in 7 family practices and one out-of-hour urgent care practice in Germany. In both settings, we included only patients with acute complaints. While this applies to most patients in out-of-hour urgent care practices, in family practices patients who came for consultations concerning chronic conditions or preventive medical check-ups could not take part.

Inclusion and exclusion criteria can be found in the study protocol.[16] Study participants were recruited by trained study nurses in waiting areas between November 22, 2021, and January 12, 2022. Recruitment was paused on and adjacent to public holidays (December 22 through December 26, December 31 through January 02).

Waiting patients were approached and informed about the study (Figure 1). Those with expressed interest were provided with detailed information about the study procedure and data privacy regulations. If they wanted to take part, they were asked to sign a written informed consent form. Once enrolled, patients received a tablet and used the app without further instructions in the waiting room. After completing the app, patients were interviewed by a study nurse in a separate room with an identical procedure.

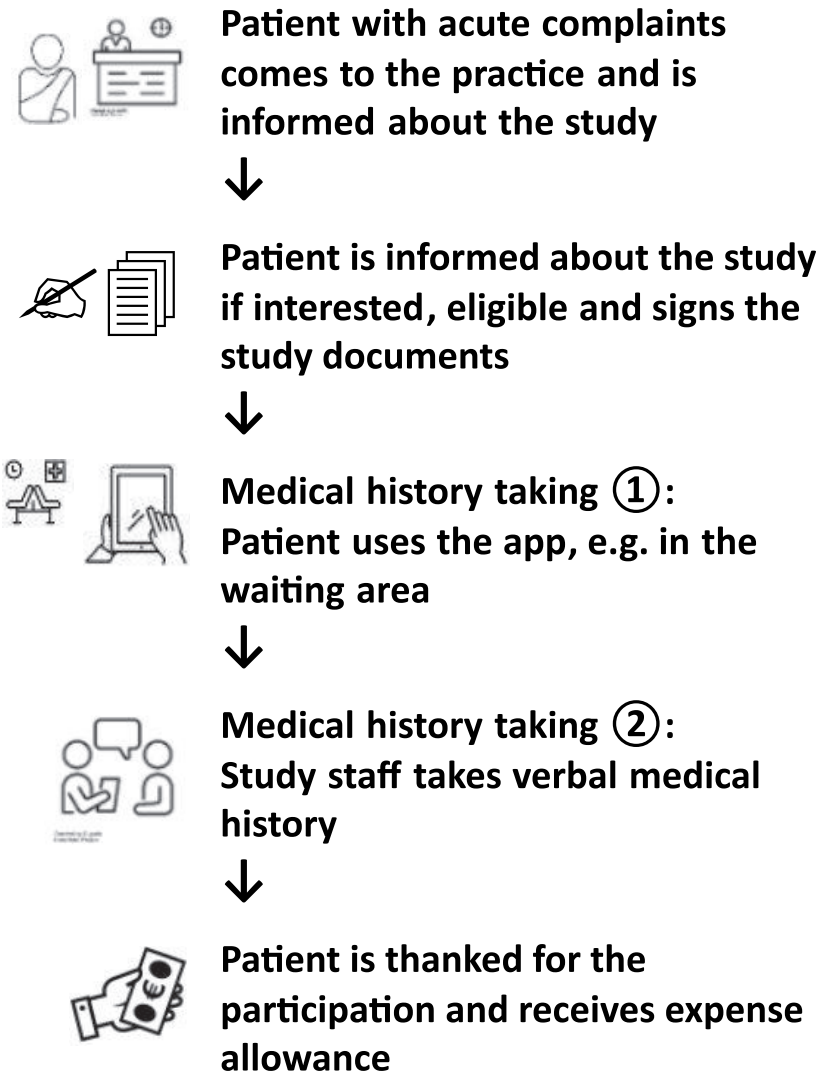
Collected patient data

Collected data included patients' age, sex assigned at birth, weight, and height both through the app-based and in-person interviews. Body mass index (BMI) was calculated and categorized using the cut-offs for adults introduced by the World Health Organization: <18.5, 18.5-24.9 kg/m^2, 25.0-29.9, and ≥ 30.0 kg/m^2 as underweight, normal, overweight, obese, respectively.[17] Patients were asked to rate the severity of their complaints on a 5-point scale. Information on whether the patient is a native German speaker was collected in the app-based interview. Patients' health literacy was assessed with 3 items focusing on understanding the physician, searching and evaluating health information that were derived from the European Health Literacy Survey adapted for the German language.[18,19]

Patients who selected "male" as their sex could choose among 36 complaints, those who selected "female" could choose 2 additional complaints: "unusual vaginal discharge" and "unusual vaginal bleeding." To facilitate finding the appropriate complaint, complaints were presented in groups based primarily on body location.

Outcomes

The primary endpoint was the concordance operationalized as exact matches between the data collected in the app-based interview and the in-person interview. The in-person interview was used as the reference. For numerical (weight, height, and age) and ordinal data (perceived severity of complaints), correlation coefficients were used as outcome. Concordance between multiple complaints was defined as an exact match of all 38 complaints between both collection methods. In addition, we calculated how many patients reported additional complaints (overreporting) or fewer complaints (underreporting) in the app-based interview compared to the in-person interview. For each individual complaint, the following proportions were calculated: proportion of concordant responses, correctly reported complaints in both interviews among reported complaints in the in-person interview (sensitivity), correctly unreported complaints in both interviews among unreported complaints in the in-person interview (specificity), and the negative predictive value (NPV) defined as the proportion that the disease is not present when the test is negative. Cohen's kappa quantifies the degree of agreement between 2 observers or data collection methods and was calculated for each individual complaint.[20] Kappa values <0, 0-0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80, and >0.80 are interpreted as slight agreement, fair agreement, moderate agreement, substantial agreement, and almost perfect agreement, respectively.[20] Individual complaints were categorized into somatic and mental. Mental complaints include anxiety, depression, and sleep disturbance/fatigue. All other complaints were categorized as somatic.



The figure was created with PowerPoint by Microsoft Office Professional Plus 2016, partly using icons from Noun Project (https://thenounproject.compatient by Minh DoVN, Reception by ibrandify PK, waiting room by Aficons ID, iPad by Andrey Vasiliev RU, Interview by Cuputo ID, euro payment by bfarias CL)

Figure 1. Procedure of the data collection.

Statistical analysis

Anonymized datasets were available for each patient as matched pairs, one from the app-based interview and the other from the in-person interview. Patients were excluded from statistical analysis if they did not complete both sessions.

Patient characteristics based on the data obtained during the in-person interview were reported as the number of patients and their proportions for categorical variables and means with standard deviations for continuous variables. The Pearson and Spearman correlation coefficient with 95% CIs were calculated for numeric and ordinal scaled variables, respectively. The concordance probabilities and their 95% CIs were estimated as the proportion of identical pairs when examining reported complaints between the 2 collection methods. In cases with more than 1 item, all items had to match to be considered concordant. CIs were calculated using the Clopper-Pearson interval for dichotomous variables.[21] CIs for Cohen's kappa were calculated in accordance with Fleiss et al.[22]

Two binomial logistic regressions were calculated. One regression was used as a dependent variable having at least 1 mismatch of complaints (nonconformity) between app-based and in-person interviews. The other regression used reporting less complaints (underreporting) in the app-based compared to in-person interviews as a dependent variable. We focused on underreporting because if complaints are not mentioned in the app, this omitted information may lead to diagnostic ambiguity and therefore influence treatment decisions. Over-reporting would not normally be a concern, as the complaints reported in the app can be discussed during the medical consultation. Gender, age, German as a native language, health literacy, perceived severity of complaints, and the number of reported complaints during the in-person interview were used as independent variables. Odds ratios (ORs) were reported with its 95% CI and *P*-value. An OR >1 indicates a higher

chance of nonconformity of complaints between app-based and in-person interview (model 1) or a higher chance of underreporting complaints in the app-based compared to the in-person interview (model 2). A *P*-value of <.05 was considered statistically significant. Statistical analysis was performed using R (version 4.1.3).

Results

Sample description

From 453 approached patients, 401 patients were recruited and 399 patients completed both, the app-based and in-person interviews. Characteristics of included patients can be found in Table 1.

Concordance of patients' characteristics

In total, 386 out of 399 (96.7%) patients gave the same age in the app-based and in-person interviews. A younger age in the app was reported by $n = 7$ (1.8%) patients and an older age by 1 patient (0.3%) (Figure 2). An almost perfect correlation was observed between the 2 data collection methods ($r = 0.999$ [0.999-1.000], $P \leq .001$).

In the app-based interview, 5 patients stated their age in weeks (values from 9 to 46). When the unit of measurement is changed to years, the age value for 3 of the 5 patients matches the age with the in-person interview.

Three hundred sixty-two out of 399 patients (90.7%) gave matching weight data in both interviews. Nineteen patients (4.8%) gave higher weight data (mean +3.2 kg), while 18 patients (4.5%) gave a lower weight (mean −3.6 kg) in the app-based interview (Figure 2). A strong positive correlation was observed between the 2 collection methods ($r = 0.996$ [0.996-0.997], $P \leq 0.001$). For 373 (93.5%) patients analyzed, the height information from the app-based interview and in-person interview matched. Sixteen patients (3.0%) reported a higher height in the app than in the interview, while 10 (2.5%) people reported a lower height. A high positive correlation was observed between the 2 data collection methods ($r = 0.996$ [0.996-0.997], $P \leq .001$). Therefore, the categorized BMI of 7 patients (1.8%) differed between the 2 data collection methods: 3 patients were categorized as normal instead of overweight, 2 patients were categorized as overweight instead of normal weight, and another 2 patients were categorized as overweight instead of obese in the app-based interviews compared to the in-person interview.

Table 1. Patient characteristics of patients.

Sex, *n* (%)	
Male	176 (44.1)
Female	223 (55.9)
Age (years), mean (SD)	38.4 (16.4)
Age categories, *n* (%)	
18-<30	156 (39.1)
30-64	216 (54.1)
65+	27 (6.8)
German native language, *n* (%)[a]	
Yes	328 (82.2)
No	67 (16.8)
Perceived severity of complaints, *n* (%)	
I don't feel sick	36 (9.0)
Just a little	74 (18.5)
Fairly	212 (53.1)
Very	68 (17.0)
Unbearably	8 (2.0)
Health literacy, mean (SD)[a,b]	
Understanding the physician	2.13 (0.57)
Search and understand health information	2.10 (0.72)
Evaluate health information	1.39 (0.78)

[a] Missing $n = 4$.
[b] Measured on a 4-point (0-3) Likert-scale (higher scores indicate higher health literacy levels).

Concordance of perceived severity of complaints

The majority ($n = 304$; 76.8%) of patients reported the same perceived severity of complaints in both surveys (Table 2). The perceived severity of complaints was rated higher by 11.6% ($n = 46$) of patients in the app compared to the in-person interview (red cells) and underestimated by 11.6% ($n = 46$). In 2.3% ($n = 9$) of patients, a deviation of more than 1 level was observed. A high correlation was found between data from the app-based interview and the in-person interview ($r[396] = 0.781$ [0.740-0.817], $P \leq .001$).

Concordance of stated complaints

Patients reported on average 1.73 (SD 1.38) and 1.40 (SD 0.73) complaints using the app-based and in-person interview, respectively ($P = .02$). The median number of complaints reported by patients in both interviews was 1.0 (IQR 1.0). Of the 399 analyzed patients, $n = 257$ (64.4%) gave completely identical responses throughout all 38 complaints in both data collection methods (Figure 3). Number of mismatches and the corresponding patient count are presented in Table 3. Additional complaints when using the app (overreporting) were reported by 74 patients (18.5%), and 68 patients (17.0%) did not report all of the stated complaints in the in-person interview when using the app (underreporting).

When categorizing the complaints into somatic complaints and mental health complaints, the concordance in these 2 categories was 0.67 and 0.93, respectively. The concordance of the individual complaints varied between 0.95 and 1.0 (see Table 3). Three complaints (*anal problems*, *fever*, and *insect bite*) were stated identically in both interviews by all patients. The lowest concordance was observed in the complaints *joint and muscle problems*, headache, and *sleep disturbance or fatigue*. The sensitivity varied among all individual complaints between 0.60 (leg pain/swelling) and 1.00 (*urinary discomfort*, *anus problems*, *fever*, *insect bites or sting*, *paralyses*, *injuries*, and *altered stool*). Regarding the specificity, the lowest values were observed in the complaints *sleep disturbances* and *fatigue* and the highest values in the complaints *anal problems*, *fever*, *insect bites and string*, and *groin problems* (Table 3). Cohen's kappa showed an almost perfect agreement in 13 complaints (34.25) and a substantial agreement in another 13 (34.2%) complaints between both collection methods. The NPV ranged from 0.95 to 1.00 across the complaints with the lowest value of 0.95 in the complaint *joint and muscle problems* (Table S1).

Factors associated with mismatching complaints

The first regression model revealed significantly increasing odds of nonconformity in complaints selection with an increasing number of stated complaints in the in-person interview (OR = 3.24 [2.29-4.71] $P < .001$). The second regression model showed a significant association between the factors sex, age, and the number of stated complaints in the in-person interview with underreporting complaints in the app-based interview. The male sex (OR = 1.90; CI [1.05-3.46] $P = .03$), a higher age (OR = 1.02; CI [1.01-1.04] $P = .01$), and more stated

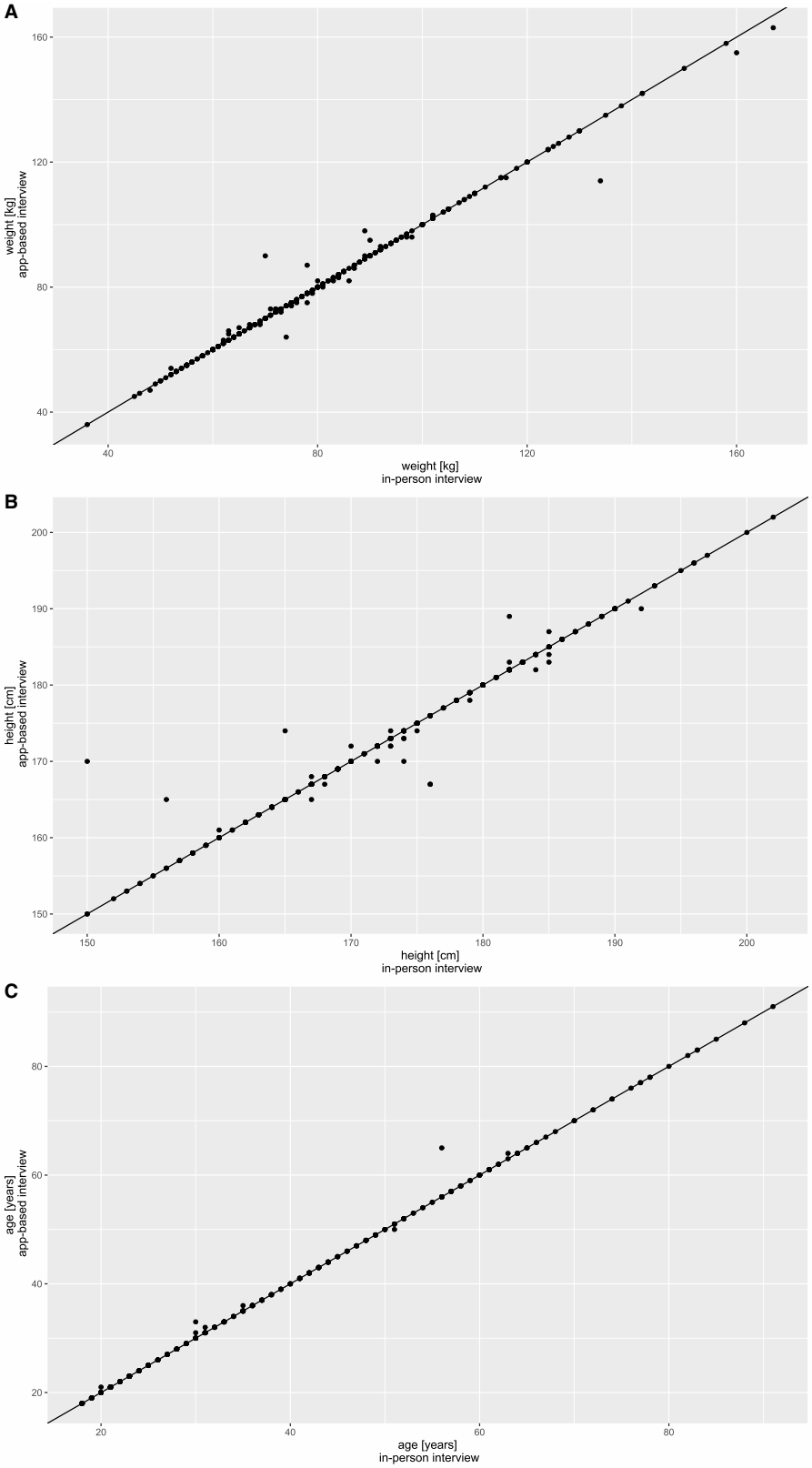


Figure 2. Stated weight (A), height (B), and age (C) in both interviews.

complaints during the in-person interview (OR = 3.14; CI [2.19-4.61] $P < .001$) were associated with increasing odds for underreporting complaints in the app-based interview (Table 4).

Discussion

In this study, we investigated the concordance of medical data in 399 patients in family practice using a tablet app

Table 2. Perceived severity of complaints.

How sick do you feel ($n = 396$)?[a]		In-person interview				
		I don't feel sick	Just a little	Fairly	Very	Unbearable
App-based consultation	I don't feel sick	24 (6.1)	5 (1.3)	2 (0.5)	1 (0.3)	0 (0.0)
	Just a little	9 (2.3)	49 (12.4)	12 (3.0)	1 (0.3)	0 (0.0)
	Fairly	3 (0.8)	19 (4.8)	183 (46.2)	21 (5.3)	2 (0.5)
	Very	0 (0.0)	0 (0.0)	14 (3.5)	44 (11.1)	2 (0.5)
	Unbearable	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	4 (1.0)

[a] Three patients did not respond during both time points; data is count (proportion).

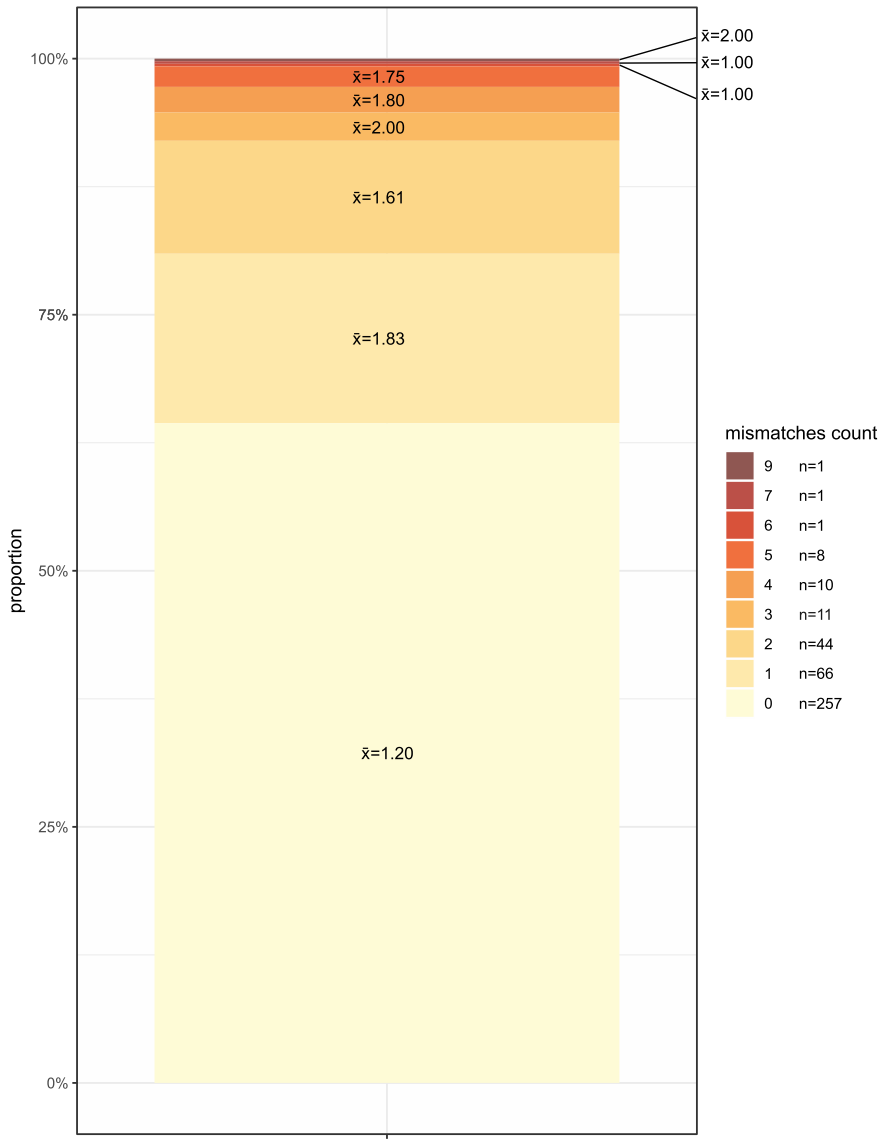


Figure 3. Number of mismatches across all 38 complaints between both collection methods x̄: mean number of stated complaints during the in-person interview stratified after mismatches count.

compared with an in-person interview. Data on patient characteristics and perceived severity of complaints was mostly concordant. Regarding the reported complaints, a concordance of 64%, 19% overreporting, and 17% underreporting across all 38 complaints was observed when using the app compared to the in-person interview. Patients who reported multiple complaints in the in-person interview had higher odds of mismatching data between the app-based data and the in-person interview. Participants who underreported complaints were more likely to be male, older, and to have stated more complaints in the in-person interview.

In contrast to previous studies,[8,9] not only the agreement between the self-reported and the queried data was examined but also the proportion of those who reported additional

Table 3. Concordance, sensitivity, specificity, and Cohen's kappa for each complaint (app-based vs in-person).

Complaint	Concordance[a]	Sensitivity (*n*/*N*)	Specificity (*n*/*N*)	Cohen's kappa[a]
Head	0.92 (0.89-0.94)	—	—	—
Eye problems	0.99 (0.97-0.99)	0.83 (5/6)	0.99 (388/393)	0.62 (0.34-0.90)
Headache	0.95 (0.92-0.97)	0.90 (38/42)	0.95 (340/357)	0.75 (0.65-0.85)
Ear problems	0.98 (0.96-0.99)	0.70 (7/10)	0.98 (383/389)	0.60 (0.36-0.84)
Dizziness	0.97 (0.94-0.98)	0.92 (12/13)	0.97 (374/386)	0.63 (0.45-0.82)
Throat, breathing and common cold	0.91 (0.88-0.94)	—	—	—
Shortness of breath	0.98 (0.96-0.99)	0.89 (8/9)	0.98 (384/390)	0.69 (0.47-0.91)
Cold or flu	0.98 (0.96-0.99)	0.88 (29/33)	0.99 (363/366)	0.88 (0.80-0.97)
Sore throat	0.98 (0.96-0.99)	1.00 (29/29)	0.98 (361/370)	0.85 (0.76-0.95)
Cough	0.97 (0.95-0.99)	0.91 (21/23)	0.98 (368/376)	0.79 (0.67-0.92)
Difficulty swallowing	0.96 (0.94-0.98)	0.86 (12/14)	0.97 (372/385)	0.60 (0.41-0.78)
Heart, chest and back	0.93 (0.91: 0.96)	—	—	—
Chest and heart problems	0.98 (0.96-0.99)	0.87 (26/30)	0.99 (364/369)	0.84 (0.74-0.94)
Palpitations	0.99 (0.97-0.99)	0.88 (7/8)	0.99 (386/391)	0.69 (0.46-0.93)
Back pain	0.96 (0.94-0.98)	0.92 (54/59)	0.97 (330/340)	0.86 (0.78-0.93)
Abdomen, digestion and urination	0.92 (0.89-0.94)	—	—	—
Abdominal discomfort	0.96 (0.94-0.98)	0.89 (50/56)	0.97 (334/343)	0.85 (0.77-0.92)
Urinary discomfort	0.998 (0.99-1.00)	1.00 (22/22)	0.997 (376/377)	0.98 (0.93-1.00)
Diarrhea	0.98 (0.96-0.99)	0.67 (6/9)	0.99 (385/390)	0.72 (0.33-0.85)
Vomiting/nausea	0.97 (0.95-0.99)	0.67 (8/12)	0.98 (381/387)	0.75 (0.38-0.83)
Changes in bowel movements/constipation	0.99 (0.97-0.995)	1.00 (5/5)	0.99 (389/394)	0.66 (0.38-0.94)
Lower body and groin	0.96 (0.94-0.98)	—	—	—
Anus problems	1.00 (0.99-1.00)	1.00 (1/1)	1.00 (398/398)	1.00 (1.00-1.00)
Unusual vaginal discharge	0.997 (0.99-1.00)	– (0)	0.997 (398/399)	—
Unusual vaginal bleeding	0.99	– (0)	0.992 (396/399)	—
Groin problems	0.998 (0.99-1)	0.83 (5/6)	1.00 (393/393)	0.91 (0.73-1.00)
Leg pain/leg swelling	0.97 (0.95-0.99)	0.60 (3/5)	0.98 (386/394)	0.36 (0.06-0.67)
General condition and psyche	0.90 (0.87-0.93)	—	—	—
Anxiety	0.97 (0.94-0.98)	0.78 (7/9)	0.97 (379/390)	0.50 (0.27-0.74)
Depression	0.99 (0.98-1)	0.95 (20/21)	0.99 (376/378)	0.93 (0.84-1.00)
Fever	1.00 (0.99-1)	1.00 (2/2)	1.00 (397/397)	1.00 (1.00-1.00)
Weight changes	0.99 (0.98-1)	1.00 (1/1)	0.99 (394/398)	0.33 (-0.15-0.82)
Sleep disturbances or fatigue	0.95 (0.92-0.97)	0.71 (15/21)	0.96 (364/378)	0.57 (0.41-0.74)
Weaknesses	0.96 (0.93-0.98)	0.71 (5/7)	0.96 (377/392)	0.35 (0.12-0.59)
Sweating	0.985 (0.97-0.99)	0.00 (0/1)	0.99 (393/398)	−0.00 (-0.01-0.00)
Forgetfulness	0.99 (0.97-1.00)	– (0/0)	0.99 (395/399)	—
Other problems	0.92 (0.89-0.94)			
Bleeding	0.99 (0.98-1.00)	– (0/0)	0.99 (397/399)	—
Joint and muscle problems	0.95 (0.92-0.97)	0.71 (40/56)	0.99 (339/343)	0.77 (0.68-0.87)
Insect bites or string	1.00 (0.99-1)	1.00 (1/1)	1.00 (398/398)	1.00 (1.00-1.00)
Paralysis/sensory problems	0.99 (0.98-1.00)	1.00 (2/2)	0.99 (394/397)	0.57 (0.13-1.00)
Fainting or blacking out	0.9975 (0.99-1)	– (0/0)	0.998 (398/399)	—
Swelling	0.98 (0.96-0.99)	0.79 (11/14)	0.99 (381/385)	0.75 (0.57-0.93)
Changes in skin, hair, or nails	0.99 (0.97-0.995)	0.88 (21/24)	0.99 (373/375)	0.89 (0.79-0.98)
Injury	0.99 (0.98-0.998)	1.00 (8/8)	0.99 (388/391)	0.84 (0.66-1.00)
Mental complaints	0.93 (0.89-0.95)	—	—	—
Somatic complaints	0.67 (0.62-0.72)	—	—	—
All complaints	0.64 (0.59-0.69)	—	—	—
All complaints including overreporting[b]	0.83 (0.79-0.87)	—	—	—

[a] With the corresponding 95% CI.
[b] Concordance including patients with exact matches and overreport in the app-based interview.

complaints in the app than in the interview (overreporting). The use of Cohen's kappa allows interpretation of the concordance of individual items but not the impact on medical relevance.[9] Kehoe et al[10] examined the sensitivity and specificity of self-reported preexisting conditions and medication compared to physician documentation. Sensitivity and specificity were lower than in our study (eg, 75% and 66%, respectively, for the disease arthritis).[10]

The frequency of complaints recorded in this study is partly comparable in frequency to previous studies in German family practices. "Back pain," the most common complaint in this study, is also one of the most common in other studies.[23,24] However, cold symptoms, which would be expected to be the most frequent due to data collection during the autumn and winter months, are only found in the fifth place, which may be explained by pandemic regulations, such as (partly) newly introduced "infection disease consultation hours" in practices[25,26] and changes in healthcare utilization during the COVID-19 pandemic.[27]

Our study found a discrepancy between self-reported information and data collected in-personal interviews including mental health complaints. One study showed that patients gave a more negative estimation of their mental health in self-administered questionnaires than in a telephone interview.

Table 4. Results of a logistic regression predicting at least 1 mismatch of main complaints and reporting less complaints in app-based compared to in-person interview.

	Model 1 (mismatching)		Model 2 (underreporting)	
Variable	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Sex				
Female	– (ref)	– (ref)	– (ref)	– (ref)
Male	1.04 (0.66-1.65)	.85	1.90 (1.05-3.46)	.03
Age				
Per year	1.00 (0.99-1.01)	.99	1.02 (1.01-1.04)	.01
German native language				
Yes	– (ref)	– (ref)	– (ref)	– (ref)
No	1.16 (0.64-2.14)	.64	1.02 (0.49-2.30)	.95
Health Literacy (scale 0-3)				
Understanding the physician[a]	1.25 (0.83-1.91)	.29	1.05 (0.61-1.82)	.86
Search and understand health information[a]	0.89 (0.63-1.27)	.53	0.83 (0.53-1.29)	.39
Evaluate health information[a]	1.06 (0.76-1.46)	.74	1.05 (0.69-1.60)	.83
Perceived severity of complaints (scale 1-5)[b]	1.22 (0.94-1.60)	.14	1.17 (0.83-1.67)	.38
Stated complaints during in-person interview[c]	3.24 (2.29-4.71)	<0.001	3.14 (2.19-4.61)	<.001

[a] Higher scores indicate higher level of health literacy.
[b] Higher scores indicate worse severity of complaints.
[c] OR increase per stated complaint; OR >1 indicates a higher chance of having at least 1 mismatch between app-based and in-person interviews (model 1) or indicates a higher chance of less stated complaints during the app-based compared to the in-person interviews (model 2).
Abbreviation: OR = odds ratio.

The authors assume that in in-personal interviews patients minimize mental health problems due to respective stigma.[28] In this study, a higher prevalence of psychological complaints was not observed in the app-based interview compared to the in-person interview. It would be interesting to investigate in further studies whether a difference could be seen in standardized questionnaires to assess symptom severity.

Other studies on medical history-taking apps have mostly focused on user satisfaction and perceived usability.[6,29] For example, Melms et al[30] evaluated patient satisfaction with an app used in the emergency department. Usability aspects, including appropriate wording and easy navigation, are prerequisites for the successful implementation of digital innovations[31] but do not allow conclusions to be drawn about the quality of the information obtained. As these aspects may influence the quality of quality, we are investigating the usability of the app in another study.

To our knowledge, this study is the first to measure medical data using a patient-used app with an in-person interview. The app "IDANA" is designed to collect health data before consultations but does not assess the concordance of the data.[32] The app "ADA" captures medical history and uses artificial intelligence to suggest a diagnosis. In a study comprising 20 clinical vignettes, ADA showed a higher diagnostic concordance than physicians in inflammatory rheumatic diseases,[33] but concordance of patient information was not assessed. Furthermore, there are medical apps to gather medical history data in emergency settings.[30,34] The study by Katayama et al[34] is a descriptive study of the frequency of pediatric emergencies without testing the application for concordance. The comparability of these studies and apps with our study is limited because medical history taking varies between specialties in terms of medical urgency, the length of interview, and the therapeutic goal.

The DASI app is very comprehensive, as it is designed to gather medical histories in family practice encompassing a heterogeneous range of complaints. A scoping review found that current apps can support family practitioners but not entirely replace them.[35] User-friendly and properly used apps could make the work in medical practices more efficient and effective. By using the DASI app, patients and physicians are provided with the necessary information prior to the medical consultation which facilitate the optimization of practice workflows. The app could also be used in other languages and generate a report in the language of the physician to overcome language barriers.

Limitations

The COVID-19 pandemic influenced patient recruitment: patients with flu-like symptoms without a negative COVID-19 test had to wait outside the practices and were not recruited. As data were collected in the autumn and winter months and the COVID-19 pandemic changed the utilization of healthcare services,[27] there may be a bias regarding the frequency of presented disease patterns. The prevalence of the complaints may influence the outcomes. Due to a low prevalence in some complaints, the sensitivity and Cohen's kappa may be imprecise and would need a bigger sample size for more precise results. For complaints that were not selected, the sensitivity or Cohen's kappa could not be calculated.

Some patients could not take part in the study, such as blind and non-German-speaking patients, resulting in a non-representative sample. Although patients over the age of 65 are the most likely to see their general practitioner,[36,37] they make up only 6.8% of our sample. Older adult's lower digital media literacy[38] could bias our results. Simultaneously, patients with more experience using contemporary media technologies may have been more likely to participate in the study than those with less experience.

The selection of complaints may have been influenced by the following: complaints partly overlap (eg, cough, sore throat, and cold), complaints may occur simultaneously (eg, diarrhea and vomiting). In these cases, patients might select several complaints or limit their choice to one complaint—and may do differently in the in-person interview. Patients who do not find their complaints may select less appropriate

complaints. In contrast, it is possible that due to the medical knowledge of the study personnel, the complaints were selected more precisely in the in-person interview. In addition, we only compared the personal data and complaints, but not the extent to which these have an influence on the diagnosis or patient treatment. Further research is needed for this.

Conclusion

Our results suggest that patients can accurately report their medical data using the app. However, the reasons for the variety of concordance between complaints and the respective medical relevance should be investigated in further studies. These results suggest that medical history apps benefit from extensive evaluation before implementation in everyday practice, especially for the recording of acute complaints.

Acknowledgments

We are very grateful to the physicians and their practice staff who allowed data collection in their practices despite the COVID-19 pandemic and the extraordinary burden that came along for health care workers. We sincerely thank all patients for their participation in this study. We thank the team at aidminutes GmbH for their commitment and support.

Author contributions

Carla Joos (Conceptualization, Methodology, Investigation, Writing—original draft), Klara Albrink (Investigation, Writing—review and editing), Eva Hummers (Conceptualization, Writing—review and editing, Funding acquisition), Frank Müller (Conceptualization, Writing—review and editing, Funding acquisition), Kai Antweiler (Data curation, Writing—review and editing), Dominik Schröder (Methodology, Formal analysis, Writing—original draft, Visualization, Data curation), Eva Maria Noack (Conceptualization, Methodology, Writing—original draft, Supervision, Project administration)

Supplementary material

Supplementary material is available at *JAMIA Open* online.

Funding

This research was funded by the German Innovation Fund (funding number 01VSF19050) of the Federal Joint Committee (G-BA). The funders have not influenced the design of this study and did not play any role during its implementation, that is, data collection and analysis, interpretation, and publication of the results.

Conflicts of interest

None declared.

Data availability

The datasets are available are not publicly but are available from the authors upon reasonable request, under consideration of the existing ethics committee vote and the legal framework conditions.

Ethics approval and consent to participate

The Medical Ethics Committee of the University Medical Center Göttingen (Ethics Approval No. 26/3/21) approved the study. Written informed consent was collected from all patients before their inclusion in the study. Study participation was voluntary for patients. Data was collected anonymously.

References

1. Keifenheim KE, Teufel M, Ip J, et al. Teaching history taking to medical students: a systematic review. *BMC Med Educ*. 2015;15:159. https://doi.org/10.1186/s12909-015-0443-x
2. Irving G, Neves AL, Dambha-Miller H, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*. 2017;7:e017902. https://doi.org/10.1136/bmjopen-2017-017902
3. Hobbs FDR, Bankhead C, Mukhtar T, et al.; National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet*. 2016;387:2323-2330. https://doi.org/10.1016/S0140-6736(16)00620-6
4. McMahon LF, Rize K, Irby-Johnson N, et al. Designed to fail? The future of primary care. *J Gen Intern Med*. 2021;36:515-517. https://doi.org/10.1007/s11606-020-06077-6
5. Koch K, Miksch A, Schürmann C, et al. The German health care system in international comparison. *Dtsch Ärztebl Int*. 2011; https://doi.org/10.3238/arztebl.2011.0255
6. Arora S, Goldberg AD, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the emergency department. *West J Emerg Med*. 2014;15:35-40. https://doi.org/10.5811/westjem.2013.2.11498
7. Lawrence BJ, Kerr D, Pollard CM, et al. Weight bias among health care professionals: a systematic review and meta-analysis. *Obesity (Silver Spring)*. 2021;29:1802-1812. https://doi.org/10.1002/oby.23266
8. Jong KJM, Abraham-Inpijn L, Oomen HAPC, et al. Clinical relevance of a medical history in dental practice: comparison between a questionnaire and a dialogue. *Community Dent Oral Epidemiol*. 1991;19:310-311. https://doi.org/10.1111/j.1600-0528.1991.tb00175.x
9. Kelstrup AM, Juillerat P, Korzenik J. The accuracy of self-reported medical history: a preliminary analysis of the promise of internet-based research in inflammatory bowel diseases. *J Crohns Colitis*. 2014;8:349-356. https://doi.org/10.1016/j.crohns.2013.09.012
10. Kehoe R, Wu S-Y, Leske MC, et al. Comparing self-reported and physician-reported medical history. *Am J Epidemiol*. 1994;139:813-818. https://doi.org/10.1093/oxfordjournals.aje.a117078
11. Berdahl CT, Henreid AJ, Pevnick JM, et al. Digital tools designed to obtain the history of present illness from patients: Scoping review. *J Med Internet Res*. 2022;24:e36074. https://doi.org/10.2196/36074
12. Benaroia M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. *Int J Med Inform*. 2007;76:283-288. https://doi.org/10.1016/j.ijmedinf.2006.01.006
13. Zhou L, DeAlmeida D, Parmanto B. Applying a user-centered approach to building a mobile personal health record app: development and usability study. *JMIR Mhealth Uhealth*. 2019;7:e13194. https://doi.org/10.2196/13194

14. Gimpel H, Manner-Romberg T, Schmied F, et al. Understanding the evaluation of mHealth app features based on a cross-country Kano analysis. *Electron Mark*. 2021;31:765-794. https://doi.org/10.1007/s12525-020-00455-y
15. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, et al. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the Consensus-Based standards for the selection of health measurement instruments: Systematic review. *JMIR Mhealth Uhealth*. 2021;9:e15433. https://doi.org/10.2196/15433
16. Albrink K, Joos C, Schröder D, et al. Obtaining patients' medical history using a digital device prior to consultation in primary care: study protocol for a usability and validity study. *BMC Med Inform Decis Mak*. 2022;22:189. https://doi.org/10.1186/s12911-022-01928-0
17. World Health Organization. *The SuRF Report 2: Surveillance of Chronic Disease Risk Factors: Country-Level Data and Comparable Estimates*. WHO; 2005.
18. Sørensen K, Pelikan JM, Röthlin F, et al.; HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health*. 2015;25:1053-1058. https://doi.org/10.1093/eurpub/ckv043
19. Schaeffer D, Berens E-M, Gille S, et al. *Gesundheitskompetenz der Bevölkerung in Deutschland vor und während der Corona Pandemie: Ergebnisse des HLS-GER 2*. Universität Bielefeld, Interdisziplinäres Zentrum für Gesundheitskompetenzforschung; 2021.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174. https://doi.org/10.2307/2529310
21. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26:404-413. https://doi.org/10.1093/biomet/26.4.404
22. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 1st ed. Wiley; 2003.
23. Klimm H-D, Peters-Klimm F, eds. *Allgemeinmedizin: Der Mentor Für Die Facharztprüfung Und Für Die Allgemeinmedizinische Ambulante Versorgung*. 5th ed. Georg Thieme Verlag; 2017.
24. Von Der Lippe E, Krause L, Prost M, et al. Prävalenz von rücken- und nackenschmerzen in deutschland. *Ergebnisse Der Krankheitslast-Studie BURDEN 2020. Published Online First*. 2021:10. https://doi.org/10.25646/7854
25. Blackenfeld H, Kaduszkiewicz H, Kochen M, et al. *SARS-CoV-2/COVID-19- Informationen & Praxishilfen für niedergelassene Hausärztinnen und Hausärzte*. 2022.
26. Westfalen-Lippe Kassenärztliche Vereinigung, ed. *Infektionssprechstunde—Risikomanagement in Arztpraxen*. Kassenärztliche Vereinigung Westfalen-Lippe; 2022.
27. Heidemann C, Reitzle L, Schmidt C, et al. Nichtinanspruchnahme gesundheitlicher versorgungsleistungen während der COVID-19-Pamdemie: Ergebnisse der CoMoLo-Studie [published online ahead of print March 16, 2022]. *J Health Monit*. https://doi.org/10.25646/9563
28. Lungenhausen M, Lange S, Maier C, et al. Randomised controlled comparison of the health survey short form (SF-12) and the graded chronic pain scale (GCPS) in telephone interviews versus self-administered questionnaires. Are the results equivalent? *BMC Med Res Methodol*. 2007;7:50. https://doi.org/10.1186/1471-2288-7-50
29. Miller S, Gilbert S, Virani V, et al. Patients' utilization and perception of an artificial intelligence–based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. *JMIR Hum Factors*. 2020;7:e19713. https://doi.org/10.2196/19713
30. Melms L, Schaefer JR, Jerrentrup A, et al. A pilot study of patient satisfaction with a self-completed tablet-based digital questionnaire for collecting the patient's medical history in an emergency department. *BMC Health Serv Res*. 2021;21:755. https://doi.org/10.1186/s12913-021-06748-y
31. Zapata BC, Fernández-Alemán JL, Idri A, et al. Empirical studies on usability of mHealth apps: a systematic literature review. *J Med Syst*. 2015;39:1. https://doi.org/10.1007/s10916-014-0182-2
32. Idana. Idana Website. 2023. Accessed May 22, 2023https://idana.com.
33. Gräf M, Knitza J, Leipe J, et al. Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy. *Rheumatol Int*. 2022;42:2167-2176. https://doi.org/10.1007/s00296-022-05202-4
34. Katayama Y, Kiyohara K, Hirose T, et al. A mobile app for self-triage for pediatric emergency patients in Japan: 4 year descriptive epidemiological study. *JMIR Pediatr Parent*. 2021;4:e27581. https://doi.org/10.2196/27581
35. Wattanapisit A, Teo CH, Wattanapisit S, et al. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. *BMC Med Inform Decis Mak*. 2020;20:5. https://doi.org/10.1186/s12911-019-1016-4
36. Bergmann E, Kalcklösch M, Tiemann F. [Public health care utilisation. Initial results of the telephone health survey 2003]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2005;48:1365-1373. https://doi.org/10.1007/s00103-005-1167-6
37. Prütz F, Rommel A, Thom J, et al. Inanspruchnahme ambulanter medizinischer leistungen in deutschland—ergebnisse der studie GEDA 2019/2020-EHIS [published online ahead of print September 15, 2021]. *J Health Monit*. https://doi.org/10.25646/8554
38. Bachman R, Hertweck F, Kamb R, et al. *Digitale Kompetenzen in Deutschland—Eine Bestandsaufnahme*. RWI—Leibniz-Institut für Wirtschaftsforschung e.V; 2021.

JAMIA Open, 2024, 7, 1–10
https://doi.org/10.1093/jamiaopen/ooae102
Research and Applications