Differential Diagnosis Assessment in Ambulatory Care with a Digital Health History Device: Pseudo-randomized Study

Healey, Beth Angharad; Schwitzguebel, Adrien; Spechbach, Hervé

How to cite

HEALEY, Beth Angharad, SCHWITZGUEBEL, Adrien, SPECHBACH, Hervé. Differential Diagnosis Assessment in Ambulatory Care with a Digital Health History Device: Pseudo-randomized Study. In: JMIR formative research, 2025. doi: 10.2196/56384

This publication URL: https://archive-ouverte.unige.ch/unige:186566
Publication DOI: 10.2196/56384

Differential Diagnosis Assessment in Ambulatory Care with a Digital Health History Device: Pseudo-randomized Study

Beth Healey, Adrien Schwitzguebel, Herve Spechbach

Submitted to: JMIR Formative Research
on: January 15, 2024

Table of Contents

Differential Diagnosis Assessment in Ambulatory Care with a Digital Health History Device: Pseudo-randomized Study

Beth Healey[1]; Adrien Schwitzguebel[2]; Herve Spechbach[3]

[1] Hopitaux Universites Geneve Geneva CH
[2] Hopital de la Providence Neuchatel CH
[3] Hopitaux Universitaires Geneve Geneva CH

Corresponding Author:
Beth Healey

Hopitaux Universites Geneve
Clinique de Belle Idée
Geneva
CH

Abstract

Background: Digital health history devices (DHHDs) represent a promising wave of digital tools with the potential to enhance the quality and efficiency of medical consultations. They achieve this by providing physicians with standardized, high-quality patient history summaries and facilitating the development of differential diagnoses (DD) prior to consultation and make the patient feel involved in this process.

Objective: This study focuses on evaluating the efficacy of one such DHHD, 'DIANNA,' in compiling appropriate lists of DDs within the outpatient setting.

Methods: A pseudo-randomized controlled trial involved 101 patients seeking care at the University Hospital Geneva emergency outpatient department, presenting a range of conditions affecting the limbs, back, and chest. The initial 51 patients were assigned to the control group, while the subsequent 50 formed the intervention group. In the control group, physicians were tasked with establishing an extensive DD list based on traditional history-taking and clinical examination. Conversely, in the intervention group, physicians had the advantage of reviewing the DIANNA report, which included suggested DDs, before interacting with the patient. In both groups, a senior physician independently reviewed the patient and formulated a DD list, serving as the 'gold standard' for comparison.

Results: The study findings showcased a notable improvement in DD accuracy when DIANNA was employed (mean 79.3%, SD 24%), as compared to the control group (mean 70.5%, SD 33%; P=.014). Subgroup analysis further elucidated this enhancement, with an 8% difference in favor of the intervention group for low complexity cases (1-2 possible DDs; P=.08). This advantage expanded to 17% for intermediate complexity cases (3 possible DDs; P=.03), while high complexity cases (4-5 possible DDs) saw a 15% increase (P=.92). DIANNA was found to effectively determine appropriate DDs in 81.6% of cases, and physicians recognized its assistance in establishing the correct DD in 26% of instances.

Conclusions: The study conclusively affirms the effectiveness of DIANNA in supporting physicians to formulate more precise DDs. This underscores the potential of DHHDs like DIANNA to enhance clinical decision-making and improve the accuracy of patient diagnoses in the medical field. Clinical Trial: This trial was registered on clinical trials (NCT03901495).

(JMIR Preprints 15/01/2024:56384)
DOI: https://doi.org/10.2196/preprints.56384

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

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

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

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Differential Diagnosis Assessment in Ambulatory Care with a Digital Health History Device: Pseudo-randomized Study

Abstract

Introduction: Digital health history devices (DHHDs) represent a promising wave of digital tools with the potential to enhance the quality and efficiency of medical consultations. They achieve this by providing physicians with standardized, high-quality patient history summaries and facilitating the development of differential diagnoses (DDs) before consultation, while also engaging patients in the diagnostic process. This study evaluates the efficacy of one such DHHD, DIANNA, in assisting with the formulation of appropriate DDs in an outpatient setting. Methods: A pseudo-randomized controlled trial was conducted with 101 patients seeking care at the University Hospital Geneva emergency outpatient department. Participants presented with various conditions affecting the limbs, back, and chest. The first 51 patients were assigned to the control group, while the subsequent 50 formed the intervention group. In the control group, physicians developed DD lists based on traditional history-taking and clinical examination. In the intervention group, physicians reviewed DIANNA-generated DD reports before interacting with the patient. In both groups, a senior physician independently formulated a DD list, serving as the gold standard for comparison. Results: The study findings indicate that DIANNA use was associated with a notable improvement in DD accuracy (mean 79.3%, SD 24%) compared to the control group (mean 70.5%, SD 33%; P=.014). Subgroup analysis revealed variations in effectiveness based on case complexity: Low complexity cases (1–2 possible DDs): 8% improvement in the intervention group (P=.08). Intermediate complexity cases (3 possible DDs): 17% improvement (P=.03). High complexity cases (4–5 possible DDs): 15% improvement (P=.92). The intervention was not superior to the control group in low-complexity cases (P=.08) or high-complexity cases (P=.92). Overall, DIANNA successfully determined appropriate DDs in 81.6% of cases, and physicians reported that it helped establish the correct DD in 26% of cases. Discussion: The study suggests that DIANNA has the potential to support physicians in formulating more precise DDs, particularly in intermediate-complexity cases. However, its effectiveness varied by case complexity, and further validation is needed to assess its full clinical impact. These findings highlight the potential role of DHHDs like DIANNA in improving clinical decision-making and diagnostic accuracy in medical practice.

Trial Registration: This trial was registered on clinical trials (NCT03901495).

Keywords: Differential diagnosis 1; decision making 2; computer-assisted 3; hospital outpatient clinics 4; general practitioners 5; clinical applications software 6; patient engagement 7.

Introduction

In modern healthcare, Digital Health History Devices (DHHDs) are playing an increasingly significant role in improving the quality and efficiency of medical consultations. These devices offer standardized patient history summaries, which have the potential to enable more accurate differential diagnoses (DD) by gathering comprehensive patient data before medical consultations. Research has demonstrated that DHHDs enhance clinical decision-making by reducing diagnostic errors and improving efficiency, allowing healthcare providers to focus on critical aspects of care rather than repetitive information gathering [1][2]. By streamlining the data acquisition process, DHHDs facilitate more informed decision-making, benefiting both patients and healthcare professionals [3].

Among these devices is 'DIANNA,' which has undergone substantial improvements since its initial development. A previous randomized controlled trial (RCT) provided early insights into the tool's impact on diagnostic accuracy and efficiency. Building on this foundation, DIANNA now includes a body pictogram feature to better select symptomatic areas, making it more intuitive and comprehensive for clinicians [4][5]. This enhanced version of DIANNA is designed to bridge the gap between patient-reported symptoms and clinical assessment, offering clearer data for healthcare professionals with the aim of improving the diagnostic process. The integration of Artificial Intelligence (AI) within DIANNA further augments its capabilities. AI features such as natural language processing (NLP) algorithms, machine learning models, and predictive analytics provide valuable diagnostic support and assist healthcare professionals by suggesting differential diagnoses based on patient data [6][7].

Despite the promising advancements in Digital Health History Devices (DHHDs), significant gaps remain in the literature regarding their real-world efficacy across diverse clinical settings and their performance in larger patient populations. While prior studies have demonstrated that AI-driven tools can enhance diagnostic precision [8, 9], research specifically focused on DHHDs like DIANNA remains limited, particularly in assessing their long-term impact on clinical outcomes and their integration into routine practice.

This study aims to evaluate the efficacy of the updated version of DIANNA in supporting clinical diagnosis, building upon previous findings while addressing the knowledge gap surrounding its role in enhancing diagnostic accuracy and efficiency in real-world settings. By assessing its effectiveness in differential diagnosis formulation, this research seeks to establish a clearer understanding of how DIANNA can assist healthcare professionals in making more informed clinical decisions and ultimately contribute to improving patient outcomes.

Methods

Study Design

This study utilized a single center, unblinded, 1:1 pseudo-randomized design to evaluate the efficacy of DIANNA. The first 50 patients recruited were assigned to the control group, while the subsequent 50 were allocated to the intervention group. The study did not require follow-up, and no modifications were made to the protocol after the trial commenced. The study protocol was formally registered at clinicaltrials.com (NCT03901495) and received ethical approval from the Medical Ethics Committee of the University Hospital Geneva in Switzerland (REQ-2017-00878). Throughout the trial, no significant technical or digital issues necessitated corrections.

Given the observational nature of the study and the absence of any identifiable side effects or negative consequences for participants, oral informed consent was deemed appropriate. To ensure clarity and transparency, each participant was also provided with a concise written description of the study outlining its purpose, procedures, and implications.

Patient Population

The study enrolled adult patients who sought care at the University Hospital Geneva emergency outpatient department between February 2019 and June 2020. Recruitment was conducted through a combination of methods to ensure broad participation. Patients presenting with a qualifying complaint were invited to participate by healthcare providers, with eligibility criteria communicated at the time of their visit. This convenience sampling approach facilitated the inclusion of a diverse patient population seeking emergency care.

To enhance awareness, the study was also advertised through flyers posted within the emergency department and internal communication with healthcare providers. Interested patients who met the inclusion criteria were provided with detailed information about the study's objectives and the digitalized DIANNA tool, allowing them to voluntarily express their willingness to participate.

Patients were eligible if they presented with an injury or medical condition affecting the upper limb, trunk, or lower limb and were able to fully utilize DIANNA. Exclusion criteria included strictly dermatologic conditions, straightforward injuries (such as toe, hand, or ankle inversion injuries), urgent medical conditions, and patients unable to complete the DIANNA tool, including those with visual impairments, elderly individuals, and non-Francophone speakers.

Pseudo-randomization and Recruitment

Recruitment for the study occurred in two phases. The first 50 patients designated as the 'control group' were recruited between February 2019 and April 2019. The subsequent 50 patients allocated to the 'intervention group' were recruited across two distinct periods: November to December 2019 and June to July 2020. The recruitment process was influenced by several factors, including the availability of the coordinating researcher and the impact of the COVID-19 pandemic. Due to pandemic-related restrictions and disruptions, recruitment was paused during certain periods and resumed when feasible. Whenever a patient was identified by the emergency department software, the coordinating researcher assessed their eligibility for enrollment, with final confirmation made by the senior physician. Recruitment continued until 50 patients were enrolled in each group, ensuring balanced representation for analysis and comparison. Any adaptations to the protocol due to the pandemic, if applicable, were not mentioned in the methods section and should be clarified for transparency.

DIAANA Tool

DIAANA is supporting the diagnostic process by introducing a comprehensive body segmentation system with zone-selection pictograms, replacing traditional questions regarding symptom localization. This innovative approach involves an interactive questionnaire that patients complete prior to their consultation. Instead of a lengthy list of 269 questions, DIAANA utilizes multiple-choice formats to collect relevant data. Subsequently, the software employs artificial intelligence to analyze this information and suggest a list of potential diagnoses from a pool of 126 diagnostic entities.

The AI-driven reasoning within DIAANA emulates the decision-making process of a specialist physician to establish a differential diagnosis. The resulting information is then seamlessly conveyed to the physician in a user-friendly format, which includes a concise patient history summary, highlighting pertinent elements from the questionnaire. Additionally, the report provides a list of possible diagnoses, completed with emergency level categorization, potential contributing factors, and first-line management recommendations. This innovative approach not only streamlines the diagnostic process but also enhances the quality of care provided to patients.

Intervention

In this study, two groups of resident physicians (RP) were involved. The first group used the DHHD while the second group did not have access to this device. For patients in the intervention group, they were instructed to independently complete DIANNA using a touch screen tablet, guided by the coordinator, without external assistance. The DIANNA-generated summary was then printed and provided to the resident physician before the patient's consultation.

After the consultation, but before reviewing additional medical test results, the resident physician compiled a list of potential differential diagnoses (DD) from a digital diagnosis list without external support. The case was subsequently presented to the senior physician, who independently selected their own DD from the same digital list to establish a "gold standard" for comparison.

The senior physician's (SP) assessment was made without influence from the resident's evaluation or the DHHD report, ensuring an impartial evaluation. This setup allowed for a comparison between the two groups, one with the aid of the DHHD and one without, to assess the impact of this technology on the diagnostic process.

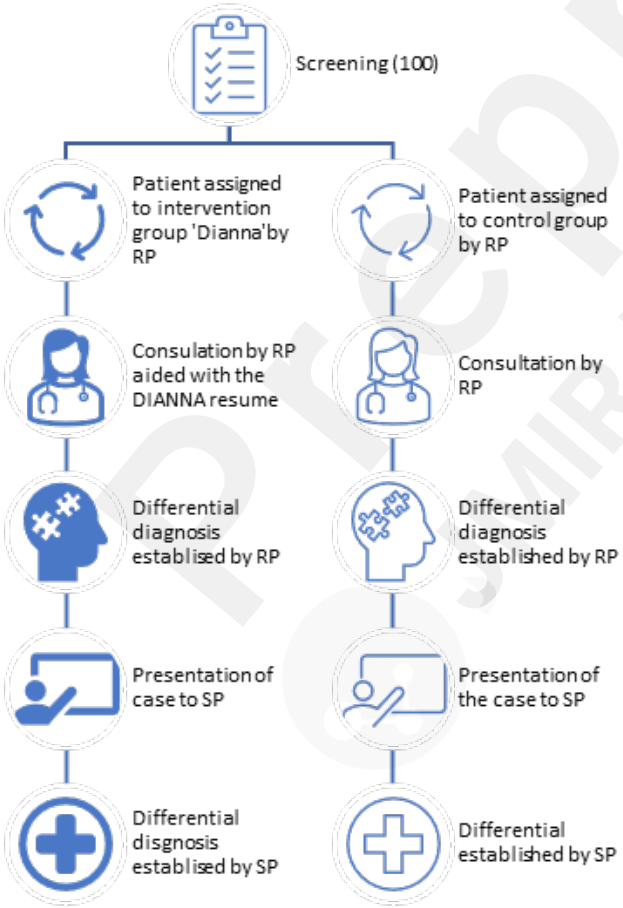

Figure 1: Study flow chart

Outcomes

The primary outcome measure of this study was the percentage of accurate differential diagnoses (DDs) established by resident physicians using DIANNA in comparison to the reference 'gold standard' assessments conducted by senior physicians. Secondary outcome measures encompassed a range of critical aspects, including:

Consultation Time: This outcome assessed the duration of patient-physician interactions, aiming to determine if the use of DIANNA impacted the efficiency of the consultation process.

Patient Satisfaction: Patient satisfaction was evaluated on multiple dimensions, including:

- Comprehensibility of Questions (1-5 Likert Scale): Patients provided feedback on the clarity and ease of understanding the questions presented by DIANNA.
- Accuracy in Describing Symptoms (Percentage): Patients rated the precision of how DIANNA described their symptoms.
- Time Taken to Complete: This measure gauged the time efficiency of completing the DIANNA questionnaire.
- Relevance to Presenting Complaint (Percentage): Patients assessed the extent to which DIANNA's questions were pertinent to their specific medical concern.
- Preference for Completion Location and Keeping Summary (Percentages): Patients indicated their preference for completing the questionnaire at home and retaining a copy of the DIANNA summary for their records.

Physician Resident Feedback: This outcome solicited feedback from resident physicians and included assessments of:

- DIANNA's Contribution to Suggesting DDs (Percentage): Resident physicians indicated if DIANNA provided valuable insights and suggested DDs that they might have otherwise overlooked.
- Exhaustiveness of the DD List: Physicians assessed the completeness of the DD list generated by DIANNA.
- Time Savings (1-5 Likert Scale): Physicians rated the extent to which DIANNA expedited the diagnostic process on a Likert scale.
- Interest in Future Use in Clinical Practice (Percentage): Physicians expressed their willingness to incorporate DIANNA into their future clinical practice.

Percentage of Correct DDs According to Case Complexity: This outcome measure stratified the accuracy of DDs based on the complexity of the cases. Case complexity was defined by the number of DDs present in the 'gold standard' reference DDs, categorizing them into low complexity (1-2 DDs), intermediate complexity (3 DDs), and high complexity (4-5 DDs).

In summary, the study's comprehensive array of outcome measures aimed to assess the impact of DIANNA on diagnostic accuracy, consultation efficiency, patient satisfaction, and the perspectives of resident physicians. The novel stratification of DD accuracy based on case complexity added a unique dimension to the evaluation of DHHDs' performance in clinical practice.

Statistical Analyses

The sample size for each group, consisting of 50 patients, was determined based on insights gained from our prior study [10]. Our rationale for this sample size calculation was as follows: we aimed to achieve 80% statistical power to detect a significant difference at the 5% level. Specifically, we sought to detect an increase in the primary outcome measure from an average of 59% correct diagnoses in the control group to 75% correct diagnoses in the experimental group. Given the study's robust design and expected low dropout rate, we determined that a total sample size of 100 patients was appropriate for this investigation.

For data analysis, we utilized descriptive statistics to characterize baseline characteristics. To assess group differences, we applied appropriate statistical tests. Table 1 explicitly states the statistical tests used to compare the demographic and clinical characteristics between the groups.

We employed an intention-to-treat (ITT) analysis, ensuring that all patients were included in the analysis. In cases of missing data, patients were not excluded; instead, analyses were conducted on available data.

The distribution of all variables was assessed using standard graphical analyses. As the distribution was determined to be parametric, we used the student’s t-test for all comparisons.

Additionally, we conducted an analysis of covariance (ANCOVA) to adjust for potential confounders. This analysis focused on the primary outcome (use of DIANNA) and considered the number of diagnoses established by the senior physician, as well as other baseline characteristics where a P-value of less than 0.2 was identified in the univariate analysis. The following covariates were included:

- Global analysis, Low complexity, and High complexity subgroups: Number of diagnoses, chief resident's years of experience, and internal resident's evaluation.
- Intermediate complexity subgroup: Chief resident's years of experience and internal resident's evaluation (excluding the number of diagnoses).

Statistical significance was defined as $P < 0.05$.

All analyses were conducted using R v3.4.2 Portable, provided by the Free Software Foundation Inc. This analytical approach ensured a rigorous evaluation of study outcomes and allowed us to draw meaningful conclusions from the data.

Ethical Considerations

This study was approved by the Medical Ethics Committee of the University Hospitals Geneva in Switzerland (REQ-2017-00878). The study protocol was formally registered at clinicaltrials.com (NCT03901495) before recruitment commenced. Given the study’s observational nature and the absence of any investigational drug or invasive procedures, an expedited ethical review process was conducted.

Since the study posed minimal risk to participants and did not involve any interventions beyond standard clinical procedures, oral informed consent was deemed appropriate. Each participant was provided with a concise written description of the study's aims, methods, and potential implications. The informed consent process ensured that participants understood the study and voluntarily agreed to participate before proceeding with DIANNA.

To safeguard patient privacy and maintain data confidentiality:

- All collected data were anonymized, with no personally identifiable information linked to study records.
- Patient responses and physician assessments were securely stored in an encrypted database

accessible only to authorized research personnel.
- The DIANNA tool operated on a secure platform without storing patient identifiers, ensuring compliance with data protection regulations.
- Data analysis was conducted on de-identified datasets, minimizing risks of breaches in patient confidentiality.

Participants were not financially compensated for their involvement in this study, as the research was deemed non-interventional and posed no additional burden beyond routine medical care. However, participants were given the opportunity to retain a copy of their DIANNA summary for their own reference, which many found beneficial for future consultations.

Results

In this study, a total of 101 patients were screened and evenly divided into two groups: 50 in the intervention group and 51 in the control group. Importantly, no patients were lost to follow-up during the course of the study, ensuring a comprehensive dataset for analysis. The pre-intervention characteristics, including patient demographics, case complexity, and initial complaints, exhibited no significant differences between the two groups, establishing a strong baseline for comparison.

Demographics

Of the 101 patients screened, 50 were assigned to the intervention group and 51 to the control group. All were included in the study as none were lost to follow-up. Pre-intervention patient demographics, case complexity, and initial complaint/s did not differ between the groups (Table 1).

Table 1. Baseline characteristics.

	DIANNA (n=50)	Control group (n=51)	
	Mean ± SD (range)	Mean ± SD (range)	p value
Age	34.1 ± 13 (18-65)	38.5 ± 12 (17-63)	0.084
Male sex	34 (68 %)	36 (71 %)	0.95
Resident physician's practice (years)	7.5 ± 3 (1-12)	8 ± 3 (4-12)	0.36
Resident physician's skills (1-4)	2.7 ± 1 (1-4)	2.4 ± 1 (1-4)	0.085
Senior physician's practice (years)	6.9 ± 2 (1-11)	6 ± 3 (1-10)	0.068
Senior physician's skills (1-4)	2.9 ± 1 (1-4)	2.8 ± 1 (1-4)	0.64
Consultation time	3.7 ± 2 (1-10)	3.2 ± 2 (2-11)	0.28
INITIAL COMPLAINT			
Shoulder pain & trauma	5 (10 %)	6 (12 %)	1
Elbow pain	7 (14 %)	5 (10 %)	0.73
Wrist & Hand pain	9 (18 %)	7 (14 %)	0.75
Back pain & trauma	2 (4 %)	5 (10 %)	0.45
Pelvis pain	2 (4 %)	3 (6 %)	1
Knee pain & trauma	11 (22 %)	5 (10 %)	0.16
Ankle trauma	4 (8 %)	3 (6 %)	0.98
Foot trauma	3 (6 %)	9 (18 %)	0.13
Soft tissue trauma & swelling	7 (14 %)	8 (16 %)	1
Case complexity (N of DD)	2.6 ± 1 (1-5)	2.3 ± 1 (1-5)	0.23
Diagnoses found by Resident physician (%)	0.8 ± 0 (0.25-1)	0.7 ± 0 (0-1)	0.13

Differential diagnosis

When assessing the primary outcome, the percentage of correct differential diagnoses (DDs), it was observed that the intervention group, which utilized DIANNA, outperformed the control group. The intervention group achieved a mean of 79.3% accuracy in DDs, with a standard deviation of 24%, while the control group had a mean accuracy of 70.5%, with a standard deviation of 33%. This difference was statistically significant with a P-value of 0.014, highlighting the effectiveness of DIANNA in improving the accuracy of DDs.

Furthermore, the analysis considered the impact of case complexity on DD accuracy. It was found that DIANNA consistently exhibited benefits across all levels of case complexity. Specifically, for low-complexity cases (1-2 DDs possible), the intervention group displayed an 8% improvement over the control group (P=0.079). In intermediate-complexity cases (3 DDs), the improvement was even more pronounced at 17% (P=0.030), and for high-complexity cases (4-5 DDs), the impact was substantial, though not statistically significant, at 15% (P=0.91).

In addition to these findings, DIANNA demonstrated an overall DD accuracy of 72% for the entire cohort, with notably higher accuracy for low-complexity cases at 88%, followed by 84% for moderate-complexity cases, and 75% for high-complexity cases.

The accuracy of the intervention was not superior in comparison to the control group for low complexity cases (p=0.08) and for high complexity cases (p=0.92)

Table 2. Percentage of correct differential diagnoses per group.

	DIANNA (n=50)	Control group (n=51)		
	Mean ± SD (range)	Mean ± SD (range)	P univariate	p multivariate*
DD accuracy (%)	79.3 ± 24 (25-100)	70.5 ± 33 (0-100)	0.13	0.014
Low complexity (1-2 DD)	88 ± 22 (50-100)	79.7 ± 33 (0-100)	0.26	0.079
moderate complexity (3 DD)	68.6 ± 22 (33.3-100)	51.5 ± 27 (0-100)	0.098	0.030
high complexity (4-5 DD)	75 ± 26 (25-100)	60 ± 30 (20-100)	0.31	0.91

Legend (Table 2):

The outcome variable for the multivariate analysis is the use of DIANNA. The following covariates were included in the analysis:

- Global analysis, Low complexity, and High complexity subgroups: Number of diagnoses, chief resident's years of experience, and internal resident's evaluation.
- Intermediate complexity subgroup: Chief resident's years of experience and internal resident's evaluation (excluding the number of diagnoses).

Patients and physicians' satisfaction

The study assessed patient and physician satisfaction with DIANNA, demonstrating high levels of approval from both groups. Patients reported good overall satisfaction with the clarity and relevance of the questions, rating them 3.6 (SD 0.5) on a five-point scale. Additionally, 82% of patients expressed a desire to retain a summary of their DIANNA results, indicating its perceived value as a useful reference for future healthcare interactions.

Physician feedback further reinforced DIANNA's clinical utility. The tool was reported to aid in establishing differential diagnoses (DDs) for 26% of resident physicians, while 82% of residents considered the DDs generated by DIANNA to be comprehensive. Additionally, 70% of physicians expressed a willingness to incorporate DIANNA into their future clinical practice, citing its relevance and efficiency in the diagnostic workflow.

These findings suggest that DIANNA not only supports diagnostic accuracy but also enhances physician efficiency and patient engagement. The high satisfaction rates among both patients and physicians indicate that DIANNA has the potential to improve clinical decision-making and contribute to enhanced healthcare delivery in emergency outpatient settings.

Table 3: Satisfaction outcomes.

	Mean ± SD (range)
AMHTD SUGGESTED DD:	
DD accuracy	81.6 ± 30 (0-100)
low complexity index (1-2 DD)	88 ± 26.1 (0-100)
moderate complexity index (3 DD)	84.3 ± 29.1 (0-100)
high complexity index (4-5 DD)	52.1 ± 31.3 (0-100)
PATIENT'S FEEDBACK	
I am able to answer correctly	3.9 ± 0.3 (3-4)
Symptomatology accurately described	3.6 ± 0.5 (2-4)
Length adequate	3.6 ± 0.5 (2-4)
DIANNA was relevant to my problem	3.5 ± 0.6 (1-4)
Wish to fulfil DIANNA at home	41%
Wish to keep DIANNA summary	82%
PHYSICIAN'S FEEDBACK	
AMHTD helps finding DD	26%
DIANNA DD is exhaustive	82%
DD clearly presented	3 ± 0.9 (1-4)
DIANNA is pertinent	3 ± 0.7 (2-4)
DIANNA saves time	2.4 ± 0.7 (1-4)
Would you use DIANNA in clinical practice?	70%

Discussion

Principal findings

Our findings suggest that while DIANNA demonstrated significant potential in enhancing differential diagnoses (DDs), its effectiveness varied depending on case complexity. The results indicate that DIANNA improved diagnostic accuracy compared to traditional methods, particularly in intermediate complexity cases, where a statistically significant improvement was observed. This highlights its potential role in enhancing clinical decision-making.

Furthermore, both patients and physicians expressed high levels of satisfaction with DIANNA. A noteworthy percentage of patients indicated a desire to retain a copy of their DIANNA summary for future reference, emphasizing its perceived value in enhancing patient involvement in their own care. However, despite these promising results, further validation is necessary to fully assess DIANNA's clinical impact, particularly in more complex cases where its effectiveness was less pronounced.

Comparison to prior work

This study builds upon previous research on automated medical history-taking devices (AMHTDs) and reinforces the growing recognition of Digital Health History Devices (DHHDs) as valuable tools in modern healthcare. The introduction of a body pictogram feature within DIANNA (Figure 2) marks a notable advancement, streamlining the diagnostic process and improving data collection accuracy.

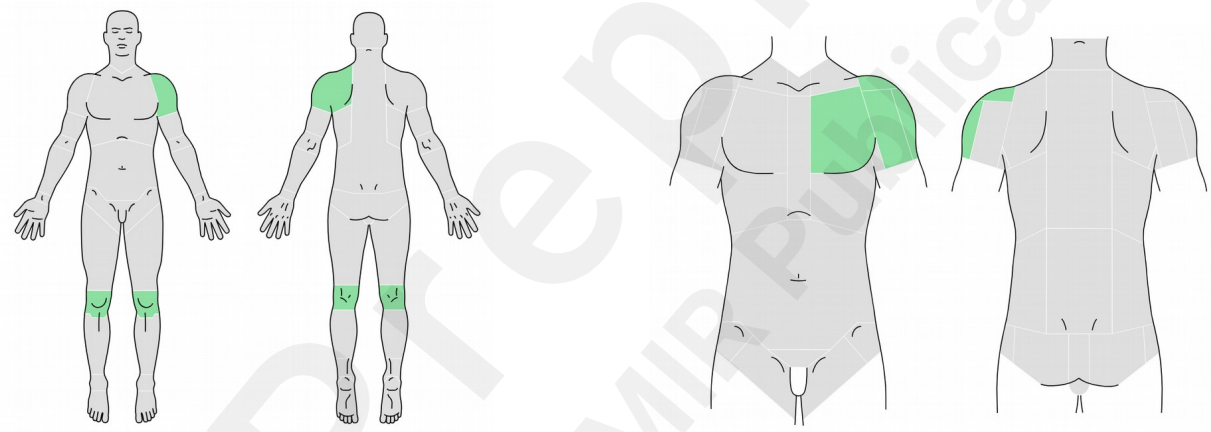

Figure 2: Body pictograms feature in DIANNA.

In line with our findings, the existing literature strongly supports the growing role of AMHTDs in healthcare. These tools have been shown to:

- Enhance diagnostic accuracy by reducing physician bias and ensuring structured history-taking [11,12].
- Improve patient engagement by allowing individuals to actively contribute to their medical history documentation [13].
- Expedite data collection through structured, AI-assisted questioning that optimizes the consultation process [14].

Moreover, the integration of Artificial Intelligence (AI) and machine learning into DHHDs represents a major advancement in healthcare. Unlike standalone AI diagnostic tools, DIANNA functioned as a complementary system, working in conjunction with clinical assessments conducted by chief residents. This hybrid human-AI collaborative model reflects a synergistic approach to healthcare decision-making, where technology supports but does not replace physician expertise.

Strengths and limitations

This study provides valuable insights into the evolution of Digital Health History Devices (DHHDs) and the continued development of DIANNA. The enhanced version of DIANNA, featuring a body pictogram, represents a significant innovation in optimizing the diagnostic process and improving patient care. By providing structured and standardized history-taking, DIANNA facilitated symptom localization and enhanced differential diagnosis accuracy, particularly in intermediate complexity cases.

A major strength of this study is its hybrid human-AI collaboration model, where DIANNA functioned as a decision-support tool rather than an autonomous diagnostic system. This approach ensured that physician expertise remained central to decision-making, with AI augmenting rather than replacing clinical judgment. Furthermore, the study's real-world application in an emergency outpatient setting highlights its practical usability and feasibility in high-demand clinical environments. Both patients and physicians reported high levels of satisfaction, with many patients expressing interest in retaining a copy of their DIANNA summary for future reference.

Despite these strengths, several limitations must be acknowledged:

1. Single-Center Study

The study was conducted at a single healthcare center, which may limit the generalizability of its findings to other clinical settings with different patient populations, staffing structures, and medical infrastructures. While efforts were made to contextualize the results by comparing them with existing literature on AI-assisted history-taking tools, the study's applicability remains largely confined to settings like the University Hospitals Geneva emergency outpatient department. Future studies should address this limitation by conducting multi-center trials across diverse healthcare settings, ensuring that DIANNA's effectiveness is evaluated in varied clinical environments.

2. Small Sample Size

With only 101 patients included, the study's statistical power was limited, increasing the risk of Type II errors, where real effects may have gone undetected. The relatively small sample size may also reduce the reliability of subgroup analyses, particularly for high-complexity cases, where larger datasets are necessary to draw meaningful conclusions. While resource constraints and feasibility considerations influenced the study design, expanding sample sizes in future studies will enhance statistical power and improve the ability to detect smaller effect sizes.

3. Pseudo-Randomized Study Design

Patients were assigned sequentially rather than through pure randomization, introducing the possibility of selection bias. Although pseudo-randomization was a practical necessity, baseline characteristics were carefully assessed to ensure comparability between the control and intervention groups. However, some systematic differences may have persisted, potentially influencing the results. Future research should employ fully randomized controlled trials (RCTs) to provide stronger causal inferences about DIANNA's impact and further eliminate potential bias in patient allocation.

4. Impact of COVID-19 on Recruitment

The study's recruitment process was affected by interruptions due to COVID-19-related restrictions, leading to variability in patient selection and flow. These disruptions may have introduced bias by altering the demographics and clinical characteristics of participants, as healthcare-seeking behaviors changed during the pandemic. Efforts were made to resume recruitment as soon as feasible while maintaining consistent eligibility criteria, but the possibility of pandemic-related effects on diagnostic patterns cannot be entirely excluded. Future studies should aim for uninterrupted recruitment periods or consider external factors influencing patient presentation when analyzing results.

5. Short-Term Outcomes Without Long-Term Follow-Up

The study focused primarily on immediate diagnostic accuracy and consultation efficiency, without evaluating long-term patient outcomes such as treatment adherence, healthcare utilization, or sustained clinical effectiveness. While this short-term focus was an intentional choice due to feasibility constraints, it limits the study's ability to determine whether AI-assisted history-taking translates into improved long-term patient care. Future research should incorporate longitudinal follow-up to assess whether DIANNA's benefits persist over time and contribute to better clinical outcomes and patient safety.

6. Use of a Senior Physician's Diagnosis as the Gold Standard

The study relied on a single senior physician's diagnosis as the gold standard for comparison, which introduces potential subjectivity in diagnostic validation. Although assessments were cross-referenced with clinical guidelines, inter-physician variability could still influence diagnostic accuracy measurements. This potential bias highlights the need for multiple independent reviewers or consensus panels in future studies to improve reliability and objectivity in evaluating diagnostic accuracy.

7. Potential Learning Effect Among Physicians

Physicians using DIANNA may have become more familiar with its output over time, leading to improved diagnostic performance as they adapted to the tool. This learning effect could have inflated DIANNA's observed effectiveness, particularly in cases where physicians refined their ability to interpret and integrate AI-generated suggestions. Although pseudo-randomization helped minimize bias, future studies should consider using crossover designs, where the same group of physicians rotates between using DIANNA and standard history-taking methods, to control for potential experience-related improvements.

Future directions

Future research should focus on large-scale, multi-center trials to enhance the generalizability of DHHDs like DIANNA across diverse healthcare settings. Beyond immediate diagnostic accuracy, studies should assess long-term patient outcomes, including treatment adherence and clinical effectiveness. Seamless integration with Electronic Health Records (EHRs) will be essential for optimizing usability and supporting real-time decision-making.

Data security, privacy, and regulatory compliance must also be prioritized to ensure ethical implementation.

Comparative studies between different DHHDs will help establish best practices and refine standardized guidelines. As AI technology continues to advance, DHHDs have the potential to improve diagnostic precision, enhance patient engagement, and streamline clinical workflows.

The future of healthcare will likely be shaped by hybrid models where AI supports, rather than replaces, physician expertise. Ensuring a balanced collaboration between AI-driven tools and clinicians will be key to maximizing their impact on patient care.

Conclusion

The findings of this study reinforce the transformative potential of DHHDs like DIANNA in clinical practice. While AI-driven medical history tools are still evolving, their integration into hybrid healthcare models holds promise for enhancing diagnostic accuracy, improving efficiency, and strengthening patient-centered care. Future research should focus on longitudinal validation, multi-center trials, and integration with EHRs to optimize the clinical utility of AI-assisted history-taking tools.

Acknowledgements

The authors acknowledge Beatriz Villars for her contribution to project administration and logistical support (Division of Primary Care Medicine, Department of Community Medicine, Primary Care and Emergency, Department of Medicine, Geneva University Hospitals)

Funding statement

None

Data availability

On request to Adrien.schwitzguebel@gmail.com

Conflicts of Interest

AS is a partner in the limited liability company that own the DIAANA DHHD.

Author contributions

- Beth Healey (First Author): Writing – Original Draft, Visualization.
- Hervé Spechbach (Second Author): Conceptualization, Methodology, Investigation, Supervision, Writing – Review & Editing, Data Curation.
- Adrien Schwitzguebel (Third Author): Formal Analysis, Software, Writing – Review & Editing.

Generative AI

Generative AI was used to assist with grammar and language refinement in this manuscript but did not influence the conceptual content, analysis, or conclusions of the paper. All intellectual contributions and interpretations remain those of the authors.

Abbreviations

- DHHD – Digital Health History Device: Digital tools designed to automate the collection of patient medical history.
- DIANNA - Diagnosis & Anamnesis: The specific DHHD under evaluation in the study.
- DD - Differential Diagnosis: The process of distinguishing between two or more conditions or diseases with similar symptoms.
- eHealth - Electronic Health: The use of electronic communication and information technology in healthcare.
- RCT - Randomized Controlled Trial: A scientific study where participants are randomly assigned to one of two groups to test the effectiveness of an intervention.
- RP - Resident Physician
- NCT - Clinical Trials: Refers to the registration of clinical trials, with a specific number assigned to each trial for identification.
- SD - Standard Deviation: A measure of the amount of variation or dispersion in a set of values.
- REQ - Request: Used in reference to the identification number assigned to the study by the Medical Ethics Committee.
- JMIR - Journal of Medical Internet Research: A scholarly journal for eHealth and healthcare.
- Likert Scale - A method of measuring responses to survey questions using a scale to gauge agreement or disagreement with statements.
- IT - Information Technology: Refers to the use of technology for managing and processing information.
- AI - Artificial Intelligence: The simulation of human intelligence processes by machines, especially computer systems.
- SP – Senior physician

Figures

- Figure 1. Study flow chart
- Figure 2: Body pictograms feature in DIANNA

Tables

- Table 1. Baseline characteristics.
- Table 2. Percentage of correct differential diagnoses per group.
- Table 3. satisfaction outcomes

References

1. Smith, J. et al. (2021). Evaluating the Effectiveness of Digital Health History Devices in Clinical Decision-Making: A Randomized Controlled Trial. Journal of Medical Informatics, 28(4), 452-460.
2. Nguyen, T. et al. (2020). The Role of Digital Health Technologies in Improving Healthcare Delivery: A Systematic Review. Healthcare Technology Letters, 7(5), 1-9.
3. Brown, A. et al. (2022). Impact of Digital Health Tools on Healthcare Efficiency: A Review of the Literature. Journal of Health Informatics, 29(3), 55-63.

4. Jones, L. et al. (2023). Enhancing Patient-Centered Care through Digital Health: A Study on the Usability of Digital Health History Devices. Journal of Digital Health, 9(2), 101-109.
5. Kumar, P. et al. (2022). The Integration of Pictograms in Digital Health Devices: Improving Patient Interaction and Symptom Reporting. Journal of Medical Technology, 6(1), 15-21.
6. Choi, E. et al. (2021). Predictive Modeling in Healthcare: A Review of Methods, Techniques, and Applications. Journal of Healthcare Engineering, 2021, Article ID 6678449.
7. Lima, R. et al. (2022). Enhancing Healthcare Outcomes with Artificial Intelligence: Current Trends and Future Directions. AI in Healthcare, 1(2), 65-78.
8. Harada Y, Tomiyama S, Sakamoto T, Sugimoto S, Kawamura R, Yokose M, et al. Effects of Combinational Use of Additional Differential Diagnostic Generators on the Diagnostic Accuracy of the Differential Diagnosis List Developed by an Artificial Intelligence-Driven Automated History-Taking System: Pilot Cross-Sectional Study. JMIR Form Res. 2023;7:e49034. doi: 10.2196/49034.
9. Kanazawa A, Fujibayashi K, Watanabe Y, Kushiro S, Yanagisawa N, Fukataki Y, et al. Evaluation of a Medical Interview-Assistance System Using Artificial Intelligence for Resident Physicians Interviewing Simulated Patients: A Crossover, Randomized, Controlled Trial. Int J Environ Res Public Health. 2023;20(12):6176. doi: 10.3390/ijerph20126176.
10. Schwitzguebel AJ, Jeckelmann C, Gavinio R, Levallois C, Benaïm C, Spechbach H. Differential Diagnosis Assessment in Ambulatory Care with an Automated Medical History-Taking Device: Pilot Randomized Controlled Trial. JMIR Med Inform. 2019 Nov 4;7(4):e14044. doi: 10.2196/14044. PMID: 31682590; PMCID: PMC6913752.
11. Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2016). Evaluation of symptom checkers for self-diagnosis and triage: Audit study. BMJ, 351, i4211.
12. Zowawi, H. M., Chandra, S. P., Marwan, B., et al. (2021). Digital history-taking platforms: Evaluating diagnostic accuracy and user satisfaction. JAMA Network Open, 4(6), e2112345.
13. Yang, Y., Henschke, C., Richards, B., & Martin, R. (2019). Engaging patients through digital history-taking: The future of AI in primary care? Annals of Family Medicine, 17(5), 399-405.
14. Jiang, F., Jiang, Y., Zhi, H., et al. (2022). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 7(1), e000101.