Translated to English from original article: Moreno Barriga E, et al. Emergencias 2017;29:391-396

Experience of Mediktor®, a new symptom checker based on Artificial Intelligence, in patients treated in an Emergency Department

Elvira Moreno Barriga[1], Irene Pueyo Ferrer[2], Miquel Sánchez Sánchez[2], Montserrat Martín Baranera[3], Josep Masip Utset[1]

A new artificial intelligence tool for assessing symptoms in patients seeking emergency department care: the Mediktor application

Objective. To analyze agreement between diagnoses issued by the Mediktor application and those of an attending physician, and to evaluate the usefulness of this application in patients who seek emergency care.

Method. Prospective observational study in a tertiary care university hospital emergency department. Patients with medical problems and surgical conditions (surgery and injuries) who did not require immediate emergency care responded to the Mediktor questions on a portable computer tablet. The software analyzed the answers and provided a list of 10 possible preliminary diagnoses in order of likelihood. The patient and the attending physician were blinded to the list to so that the usual care process would not be altered. The level of agreement between the physician's diagnosis and the Mediktor diagnosis was analyzed.

Results. A total of 1015 patients were included; 622 cases were considered valid for study. Cases were excluded if the patients did not meet the inclusion criteria, they did not have a discharge diagnosis, they had a final diagnosis expressed as a symptom or their final diagnosis was not included in the Mediktor database. The physician's diagnosis (the gold standard) coincided with one of the 10 MEDIKTOR diagnoses in 91.3% of the cases, with one of the first 3 diagnoses in 75.4%, and with the first diagnosis in 42.9%. Sensitivity was over 92% and specificity over 91% in the majority of common diagnostic groups; the κ statistic ranged from 0.24 to 0.98.

Conclusions. The Mediktor application is a reliable diagnostic aid for the most prevalent problems treated in a hospital emergency department. The general public finds it easy to use.

Keywords: Emergency department, Surgery, Injuries, Symptoms, Physician diagnosis.

Authors' affiliations
[1]Service of Intensive Care Medicine, Consorci Sanitari Integral, Barcelona, Spain.
[2]Emergency Department, Hospital Clínic, Barcelona, Spain.
[3]Department of Epidemiology and Research, Consorci Sanitari Integral, Barcelona, Spain.

Authors' contributions:
All authors have confirmed their authorship in the author responsibilities document, publication agreement and assignment of rights to EMERGENCIAS.

Corresponding author:
Elvira Moreno Barriga C/ Ricart, 16, 2º, 1º 08004 Barcelona, España.

Email:
elviramorenobarriga@hotmail.com

Article information:
Received: 4-5-2017
Accepted: 7-8-2017
Online: 9-10-2017

Editor in charge:
Agustín Julián-Jiménez, MD, PhD

Introduction

The proliferation of the internet and the easy accessibility of medical information has led to the fact that searching medical terms on the net nowadays has become a common practice in the general population[1]. If the information is not properly verified, it can lead to erroneous interpretations[2], therefore posing a risk to the user's health. In line with improving existing disinformation, sophisticated digital products called symptom checkers have proliferated[3-5] that, using computerized algorithms, interrogate the user about their symptoms, to finally offer a list of possible diagnoses, indicating with what urgency they should request medical help and to whom they should go. The true impact of the use of these symptom checkers will depend to a great extent on their degree of precision and the clinical and economic benefit they offer[6,7].

Along these lines, MEDIKTOR CORP (MC) - TECKEL SOLUTIONS, is a company whose objective is to develop and market IT solutions based on Artificial Intelligence focused on the medical field. It has developed a symptom checker called Mediktor®, aimed at facilitating and speeding up the medical diagnostic process through Artificial Intelligence algorithms. Unlike other symptom checkers that exist in the market[3-5], and because the diagnostic precision can be affected by the interaction between the user and the software, MC develops and applies to Mediktor® natural language processing algorithms, without requiring knowledge of technical terms by the user.

For the improvement of the diagnosis process, the studies that had been carried out on some of the different symptom checkers that existed in the market were made by measuring their accuracy through "virtual patient vignettes". These were obtaining an average of 58% of concordance between diagnoses, taking into account the first 20 diagnoses identified[8].

Therefore, the goal of this trial was analyzing, in a set of real patients, what is the concordance between the diagnoses issued by Mediktor® after completing the guided questioning (performs a differential diagnosis with ten possible diagnoses from greater to lower probability) with the diagnosis issued by the responsible physician (gold standard) for the same patient in the same episode and, therefore, its usefulness in patients who attend an emergency department (ED).

Methodology

Prospective observational study carried out in the Emergency Department of the Hospital Clínic de Barcelona (HCB). This is an 800-bed, third tier (high complexity) hospital, located in the city center, with a catchment area that includes approximately 550,000 inhabitants. At present, around 95,000 general emergencies are attended annually, except for Obstetrics and Gynecology, Pediatrics and Ophthalmology specialties, which are carried out in other centers. Since February 2009, the triage of patients in the ED is carried out by nursing staff, using the urgency level assignment according to the Andorran Model of Triage (MAT)[9]. The study was conducted in a 4-month period, from October 1, 2016 until January 31, 2017.

The selected patients were those to whom the triage assigned levels of MAT III, IV and V (which did not require immediate attention) according to the order of arrival and in different shifts to ensure that a representative sample was obtained. The inclusion criteria were: age > 18 years, adequate level of consciousness and ability to understand the Spanish language in order to answer the questions of the program. Patients who did not meet the inclusion criteria, those who refused to participate in the study, those who did not have a diagnosis at discharge (“USERS”), those whose final diagnosis was expressed as a symptom or sign (“NONSPECIFIC” = diagnoses not admitted in ICD-10) and cases with diagnoses not included in the Mediktor® dictionary at the time of the study (“NOT CONTEMPLATED”) were excluded.

A nurse belonging to the HCB and attached to the study (NR) interviewed the patients in the Medicine, Surgery and Orthopedic Emergency units, after having been triaged (levels III-V) and distributed following the usual procedure. The nurse introduced herself while the patients were waiting in the waiting room to be seen and explained how the study worked in an understandable language (without technical terms), confirming that the study would not change the time, diagnostic tests or treatment that may be considered for their condition. After confirming the inclusion criteria and discarding the exclusion criteria, a prudent time for reflection was established, after which written informed consent was requested for inclusion in the study and a copy of it was delivered. To the patients who agreed to participate in the study, the program assigned them a 5-digit randomized code in order to identify them. Each of the codes was generated

by the computer, with the user's data remaining completely anonymous, as initials, date of birth or national identity document number were not included. The NR registered this code in a file protected with the Medical Record (MR) number of the patient. This process ensured random coding and user anonymization. After assigning the code, the NR took the vital signs of the patient (blood pressure, heart rate, temperature and pulse oximetry) by filling in the corresponding sections in Mediktor®. Afterwards, the NR gave them the tablet so that they could enter the reason why they had come to the ED and answer themselves to the questions the program was asking them (between 7 and 20), thus carrying out the evaluation individually (self-administered). The average number of questions that the symptom checker asked among all the patients included in the study was 17 questions and the average time calculated for each session was 2 minutes and 25 seconds.

Some of the questions were answered when selected, others had a multi-response format (text, image or combination of both), and the vast majority were answered by choosing from five possible answers: YES, NO, PROBABLY YES, PROBABLY NOT, or I DO NOT KNOW. At the end of the questionnaire, the software thanked them for their participation and wished them a speedy recovery without displaying the list of diagnoses, which were concealed and recorded on the digital tablet. In this way, influencing the responsible physician, or generating conflicts if there were diagnostic differences between the physician and Mediktor®, was avoided. At the end of the questionnaire, patients were invited to answer four questions about the usefulness of Mediktor®. The questions were as follows: 1) Have you understood most of the questions well? 2) Have you found the application easy to use?, 3) Do you think that digital technologies can help improve the healthcare system?, 4) Would you like to be able to get medical guidance from your home?. All of them had NO / YES / I don't know / No answer as possible answers.

The protocol was approved by the Ethics Committee in Clinical Research of the Hospital Clínic de Barcelona.

Statistical analysis, data processing and concordance analysis were performed later, reviewing the MR of the patient and the diagnoses registered by the program, introducing them in a database. Similarity diagnostic fields were established, which allowed grouping of diagnoses expressed with different terminology.

Given that there were no precedents in the literature, a pilot study was conducted with 358 patients to estimate the sample size. After the statistical analysis and taking into account the percentage of associated losses, a minimum of 588 cases was considered necessary. The data was analyzed with the statistical program SPSS v20 (IBM®, North Castle, New York, USA), performing several calculations:

- The proportion of coincidences between Mediktor® and the attending physician as the number of coincidences over the total of valid cases. The coincidences were taken into account in the first diagnosis, in the top three diagnoses and in the top ten diagnoses provided by Mediktor®. For each of the three proportions, the 95% confidence intervals were calculated by the Clopper-Pearson exact method.

- Taking into account the most frequent diagnoses in the ED and considering the medical diagnosis at the time of discharge as a reference, the sensitivity (Sen) (probability that Mediktor® included among the first ten diagnoses, when the disease was present), specificity (Spe) (probability that the result of Mediktor® was negative when the disease was not present), positive predictive value (PPV) (probability of the disease being present when Mediktor® included it among its first ten diagnoses) and negative predictive value (NPV) (probability that the disease was not present when Mediktor® offered a negative result for that disease) were calculated, expressing each of them as percentages. The 95% confidence intervals for sensitivity and specificity were calculated using Clopper-Pearson exact formula. For the 95% confidence intervals of the predictive values, the standard logit method by Mercaldo *et al*[10] was applied. The reliability assessment methodology was performed by calculating the kappa index with specific confidence intervals for each diagnostic category.

Results

During the study period, 1,015 patients were selected according to the inclusion criteria, and a total of 622 patients were considered definitively as valid cases for the analysis. We excluded 393 patients from the study, among whom were those who were not diagnosed at discharge ("USER", 61 cases), those under 18 years ("NO 18", 21 cases), those whose final diagnosis was expressed as a symptom ("NONSPECIFIC" = diagnoses not admitted in ICD-10, 151 cases) and cases with diagnoses not included in the Mediktor® dictionary at the time ("NOT CONTEMPLATED", 160 cases).

The variability of the sample was ensured according to gender (female: 56.0%, male: 44.0%), age ranges expressed in years (18-24: 26.7%, 25-49: 52.9% , 50-75: 18.3%, and older than 75: 2.1%) and the work shifts in which the total number of cases was collected (8-14 h: 53.7%, 14-20 h: 46.3%).

The most prevalent diagnoses within the ED, taking the medical diagnosis at discharge from the ED as the standard, were defined as those groups of conditions in which there were more than 4 cases collected in our series. We can observe in each one of them the degree of agreement of the symptom checker in the ten diagnoses assigned by Mediktor®, in the first three diagnoses and in the first diagnosis (Table 1).

The coincidences, analyzing the global of included patients (622), between the medical diagnosis and the ten diagnoses of Mediktor® were in 568 cases, 91% (CI 95%: 88% to 93%); in the first three diagnoses in 469, 75% (CI 95%: 71% to 78%); and in the first diagnosis in 267 cases, 43% (CI 95%: 39% to 46%) (Table 2).

The results in terms of Sen, Spe, PPV and NPV are presented for the twelve most frequent diagnoses (Table 3), observing a Sen > 92% in 9 of the 12 groups analyzed and a Spe > 91% in 10 of them. Likewise, the kappa index calculated for the same diagnostic groups ranged between 0.24 and 0.98, obtaining the lowest value for the urinary tract condition group and the highest for the upper respiratory tract infection group (Table 4).

The degree of utility of the symptom checker was assessed with the 4 questions that Mediktor® asked at the end of the questionnaire. These questions and their results are shown in Table 5. Mediktor® is an easy-to-use program for more than 98% of respondents.

Discussion

In order to find a new tool to aid clinical diagnosis, we decided to conduct this study on Mediktor® since, in addition to being a symptom checker that was using artificial intelligence algorithms, it allowed us to perform the medical questionnaire without problems of understanding for the users since a natural vocabulary was used.

Table 1. Most prevalent diagnoses in the Emergency Department and matches between the medical diagnosis and Mediktor®

Diagnoses in the Emergency Department	Matches between medical diagnosis and Mediktor®		
	As the first diagnosis (%)	In the first 3 diagnosis s (%)	In the first 10 diagnoses (%)
Upper airway infections*	76.25	93.7	97.5
Urinary tract condition	70.6	87.2	100
Lumbociatalgia	82.2	93.3	100
Otitis	73.3	96.6	100
Contusion	76	88	100
Sprain *	79.2	87.5	91.7
Non-specific abdominal pain	4.5	45.4	100
Headache	88.9	88.88	100
Tendinitis*	81.2	81.2	81.2
Acute gastroenteritis	80	100	100
Vertigo*	84.6	84.6	84.6
Cervicalgia	53.8	61.5	100
Other diagnoses**			

* less than 100% matches is achieved when adding the cases in the three columns.

** Includes more than 200 diagnoses different where less than 4 patients were obtained included / pathology: anal fissure, constipation, whitlow nail, herpes zoster, acute diverticulitis, hemorrhoids, plantar fasciitis, mechanical costal pain, traveler's diarrhea, lower respiratory tract infection, allergic reaction, etc.

After making a comparison between the diagnoses obtained (physician and symptom checker), Mediktor® was found to have a diagnostic concordance of 91.3% of when the results were analyzed against the list of 10 possible pathologies indicated as diagnoses by Mediktor®, 75.4% with the top 3 possible diagnoses and 42.9% with the top diagnosis, with a moderate level of agreement on most analyzed patients with a kappa index ranging between 0.24 and 0.98. At the same time, the high sensitivity (>92%) and specificity (>91%) of the assessment for most of the patients analyzed could be verified.

In addition, during the realization of our study the patients felt confident using the system and had no difficulties when answering the different questions the symptom checker

was asking them. Thus, more than 98% of the respondents considered Mediktor® an easy program to use.

TABLE 2. Matches between medical diagnosis and Mediktor® according to the total of patients (622)

		CI 95%	
	% Success Mediktor	Lower bound	Upper Bound
As the first diagnosis	267 (42.9)	39.0	46.9
In the 3 first diagnoses	469 (75.4)	71.8	78.7
In the first 10 diagnostics	568 (91.3)	88.8	93.4

The medical diagnosis of the diseases suffered by our patients continues to pose difficult challenges and, although there are few studies on diagnostic accuracy, it is known that the percentage of error in medical diagnosis (including the performance of additional tests) is between 10 and 15%[11]. Arriving at a correct diagnosis is a complex process that can be influenced in a positive way by the ability to use memory to recover representative disease patterns[12-14]. Although this ability is limited since no more than seven possibilities can be managed at a time and this is subject to the possibility of clinical reasoning bias. Also, negatively, by errors related to cognitive reasoning process, as well as human factors such as fatigue, overwork and poor management of care time among others[15-16]. To date, no strategy, cognitive[17] or systemic[18], has helped to improve the diagnostic accuracy and therefore to reduce the diagnostic error.

Therefore, we consider the results of our trial to provide relevant data that lead us to consider Mediktor® to be a computer-based diagnostic aid tool that may be complementary to regular clinical practice and that its implementation in hospital EDs may accelerate the diagnostic process with an improvement in efficiency.

In addition, and after the analysis of the results, we consider future strategies with respect to the symptom checker, since Mediktor® could become a complement to conventional triage, as advanced nursing triage[19]. In this way it could be an aid to streamline diagnostic tests requests or start of treatments, therefore shortening the stay in the ED[18]. In turn, it could be considered whether Mediktor® could become a new triage system, which would require demonstrating concordance with the level of urgency against the results of currently validated systems. Other artificial intelligence algorithms in the literature have tried unsuccessfully[8].

Our study has a number of limitations.The first one, significant, is that our study has been carried out in a single location. Besides, the patients included in the study belonged to the III-V groups in MAT (triage system) and, therefore, did not require immediate medical attention in relation to severity. In addition, very few patients older than 75 years old have been included, which does not allow to reach any conclusions for this age group. In summary, it may be said that Mediktor® has been assessed only in "low risk young" patients. Also, a sizeable number of patients were excluded as “nonspecific” due to an absence of ICD-10 diagnosis at time of discharge (14%) or “not contemplated” because the diagnosis was not available in the Mediktor® database which, at the time of the trial, included 560 conditions. In any event, no study to date had analyzed the concordance between physician diagnoses and symptom checkers with real patients.

TABLE 3. Most frequent diagnoses in the Emergency Department: Sensitivity, Specificity, Positive Predictive Value and Negative Predictive Value achieved by Mediktor®.

	Sensitivity (CI 95) %	Specificity (CI 95) %	PPV (CI 95)%	NPV (CI 95) %
Upper Airway Infections	97,5 (91,3-99,7)	100,0 (99,3-100,0)	100	99,6 (98,6-99,9)
Urinary tract conditions	92,6 (82,1-97,9)	68,5 (64,5-72,3)	21,8 (19,5-24,4)	99,00 (97,4-99,6)
Lumbociatalgia	94,9 (82,7-99,4)	93,5 (91,2-95,3)	49,3 (41,5-57,2)	99,6 (98,6-99,9)
Otitis	93,7 (79,2-99,2)	93,7 (91,5-95,5	44,8 (36,9-52,9)	99,6 (98,6-99,9)
Non-Specific Abdominal Pain	88,0 (68,8-97,4)	98,0 (96,5-98,9)	64,7 (50,7-76,6)	99,5 (98,5-99,8)
Sprain	100,0 (84,6-100,0)	98,3 (96,9-99,2)	68,7 (54,3-80,3)	100
Contusion	100,0 (86,3-100,0)	93,1 (90,8-95,0)	37,9 (31,2-45,0)	100
Tendinitis	81,2 (54,3-95,9)	93,4 (91,1-95,2)	24,5 (18,2-32,2)	99,5 (98,5-99,8)
Headache	100,0 (80,5-100,0)	91,9 (89,4-93,9)	25,8 (20,9-31,2)	100
Acute Gastroenterittis	94,1 (71,3-99,8)	89,3 (86-5-91,6)	19,7 (15,9-24,2)	99,8 (98,8-99,9)
Cervicalgia	100,0 (71,5-100,0)	95,2 (93,2-96,8)	27,5 (21,0-35,1)	100
Vertigo	84,6 (54,5-98,1)	98,7 (97,4-99,4)	57,9 (39,9-73,9)	99,7 (98,8-99,9)

PPV: Positive Predictive Value; NPV: Negative Predictive Value; CI: Confidence Interval

TABLE 4. Most frequent diagnoses in the ED: concordance (kappa index) in diagnosis between Mediktor® and the emergency physician

Diagnoses	Kappa Index	CI 95 (%)
Upper airway infections	0.986	0.965-1.000
Urinary tract conditions	0.248	0.185-0.310
Lumbociatalgia	0.618	0.511-0.724
Otitis	0.577	0. 459-0.694
Non-specific abdominal pain	0.733	0.604-0.862
Sprain	0.807	0.690-0.923
Contusion	0.522	0.397-0.646
Tendinitis	0.351	0.207-0.496
Headache	0.383	0.252-0.513
Acute Gastroenteritis	0.295	0.180-0.409
Cervicalgia	0.415	0.247-0.584
Vertigo	0.680	0.492-0.867

CI: Confidence Interval

TABLE 5. Degree of acceptance and usefulness of Mediktor® as perceived by patients.

Questions	Answers			
	Yes (%)	No (%)	DK/NA (%)	Total (%)
1. Did you understand most of the questions easily?	98.8%	0.7%	0.4%	100%
2. Did you find the application easy to use?	98.2%	1.1%	0.7%	100%
3. Do you think information technologies can help improve the healthcare system?	95.3%	3.9%	0.9%	100%
4. Would you like to be able to get medical advice from home?	85.3%	13.8%	0. 9%	100%

DK/NA: Does not know or does not answer the question

In conclusion, our findings suggest that Mediktor® has a high degree of agreement with physician diagnosis, which makes it a useful tool to help diagnose the most prevalent diseases of an ED and easy to use by the general public. As a future research line, a trial based on analyzing the agreement between the level of urgency assigned by Mediktor® against the result of triage systems already validated to date could be considered.

Conflict of Interests

The authors declare not having any conflict of interest.

Funding

The authors declare there is no external funding for this article.

Ethical considerations

The study was approved by the Ethics and Clinical Research Committee of Hospital Clínic de Barcelona.

Informed consent was obtained from the participants. All authors have confirmed the maintenance of confidentiality and respect for patients' rights in the author's responsibilities document, publication agreement and assignment of rights to EMERGENCIAS.

Article not commissioned by the Editorial Board and with external peer review

Bibliography

1. Barnett ML, Linder JA. Antibiotic prescribing to adults with sore throat in the United States, 1997-2010. JAMA InternMed.2014;174:138-40.
2. Black P. The dangers of using Google as a diagnostic aid. Br J Nurs. 2009;18:1157.
3. North F, Ward WJ, Varkey P, et al. Should you search the Internet for information about your acute symptom? Telemed J E Health. 2012;18:213-8.
4. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res. 2014;16:e16.
5. North F, Varkey P, Laing B, et al. Are e-health web users looking for different symptom information than callers to triage centers? Telemed J E Health. 2011;17:19-24.
6. Fox S, Duggan M. Health Online 2013. Internet and American life project. PewResearch Center and California HealthCareFoundation. 2013:4.
7. Saczynski J, Yarzebski J, Lessard D, et al. Trends in pre-hospital delay in patients with acute myocardial infarction (from the Worcester Heart Attack Study). Am J Cardiol.2008;102:1591.
8. Semigran H. L ,Linder J. A, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ. 2015;351:3-5.

9. Jiménez JG. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias. Emergencias. 2003;15:165-174.
10. Mercaldo ND, Lau KF, Zhou XH.Confidence intervals for predictive values with an emphasis to case-control studies. Statistics in Medicine. 2007;26:2170-21
11. Graber M.L. The incidence of diagnostic error in Medicine. BMJ QualSaf.2013;22: ii21-ii27.
12. Alonso Carrión L, CastellsBescós E, Alba Conejo E, Ruiz Cantero A. El razonamiento clínico como fuente de error. Med Clin (Barc). 2002; 118:587-9.
13. Alonso Carrión L, Bermúdez F, Abarca M, García Almeida JM. ¿Diagnosticar mejor? Savia nueva para un árbol viejo. MedClin (Barc). 2011; 137:561-4.
14. Charlin B, Boshuizen H.P., Custers E.J., Feltovich P.J. Scripts and clinical reasoning. Med Educ. 2007;41:1178-1184.
15. Scott I. Errors in clinical reasoning: causes and remedial strategies. BMJ. 2009;339: 22-25.
16. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. AcadMed. 2003;78:775-80.
17. Graber M.L. ,Kissam S. ,Payne V.L. , Meyer A.N.D., Sorensen A.,Lenfestey N. Cognitive interventions to reduce diagnostic error: a narrative review. BMJ. QualSaf.2012;21: 535-557.
18. Singh H. ,Graber M.L. ,Kissam S.M. ,Sorensen A.V. , Lenfestey N.F. TantSystem-related interventions to reduce diagnostic errors: a narrative review BMJ QualSaf.2012;21:160-17.
19. Carballo Cardona C. Triaje avanzado: es la hora de dar un paso adelante. Emergencias. 2015;27:332-5