Implementing Medical Chatbots: An Application on Hidradenitis Suppurativa

Manuel Walss^a Florian Anzengruber^b Ahmed Arafa^{c, d} Vahid Djamei^e
Alexander A. Navarini^b

^aDepartment of Medicine, University of Zurich, Zurich, Switzerland; ^bDepartment of Dermatology, University Hospital Zurich, Zurich, Switzerland; ^cDepartment of Public Health, Faculty of Medicine, Beni-Suef University, Beni-Suef, Egypt; ^dDepartment of Public Health, Graduate School of Medicine, Osaka University, Osaka, Japan; ^eSwiss4ward, Zurich, Switzerland

Keywords

Chatbot · Digital health · Hidradenitis suppurativa · Tentative diagnosis

Abstract

Background: The use of digital health resources is growing quickly as they are easily accessible and permit self-evaluation. Yet, research on consumer health informatics platforms is insufficient. Chatbots, interactive conversational platforms based on artificial intelligence, can facilitate access to specific information. Hidradenitis suppurativa (HS) is burdensome and has a high threshold for consultation. **Objectives:** We aimed to identify the most important principles for the assembly of medical chatbots through the analysis of usage data. **Methods:** The *HS Chatbot*¹ is a question-and-answer platform in the style of a chatbot. Usage data were collected over the course of a year. 254 responses were statistically analysed. **Results:** 239 users were alleged patients. 82.9% were looking for a tentative diagnosis. The users were on average 32.49 (± 11.33) years old and predominantly female (70.2%). The average number of clicks per visit on the

website was 14.69 (± 8.83). **Conclusions:** A medical chatbot has to be customised to the specific subject whilst general principles have to be considered. High-quality information has to be available in just a few clicks. People concerned about HS are looking for a diagnosis online and often have not seen a doctor previously. Guidance towards appropriate care should be provided.

© 2021 The Author(s)
Published by S. Karger AG, Basel

Introduction

In recent years, patients have increasingly come to rely on digital sources of health information. In the field of dermatology, digital applications help raise awareness of skin conditions, offer access to information, triage and monitor patients [1, 2]. They allow clinicians to collect information on people with disfiguring conditions, hidradenitis suppurativa (HS) being one of the most important in this regard. However, patients looking for help online are met with an overwhelming amount of information of uncertain reliability. An accredited source of health knowledge is thus required to provide easily accessible, high-quality information [3, 4]. This source should

¹ www.hs-bot.com (16.06.2020).

Fig. 1. Flowchart of the Materials and Methods.

allow access to specialist-level medical information without requiring high digital health literacy. It should provide interactive communication, similar to a consultation with a doctor, in order to strengthen the user's adherence to preventive and therapeutic recommendations [5, 6].

Hidradenitis suppurativa is a multifactorial chronic inflammatory skin disease primarily affecting areas rich in apocrine glands. HS is characterised by recurrent painful nodules and abscesses that are complicated by sinuses and scars. It affects around 1% of the general population, predominantly women [7, 8]. HS was classified by Hurley into three stages of severity of symptoms. More than a third of new HS patients are already at Hurley stages II and III, potentially due to the diagnostic delay [9–11]. Such a delay may be attributed to lacking knowledge about the disease among the general population. Due to shame about a disfigured genital area, patients may not have the lesions examined by a family doctor or specialist. Meanwhile, online searches for HS have increased in number². The readability, quality and timeliness of HS-related online content is, however, questionable [12]. We therefore decided to use HS as an exemplary disease for the implementation of a chatbot in order to support specialists in their communication with HS patients, find out how HS patients are utilising health informatics and provide groundwork for creating further chatbots.

The *HS Chatbot* enables communication with a virtual database on an individual level. Using a catalogue of predefined items and responses, chatbot users were able

to choose their own path according to their interest. The set-up allowed collecting data on chatbot use, users' interests and tentative diagnoses.

Materials and Methods

To generate the data for the study, we first built a question-and-answer platform, mimicking a chatbot. The contents included three main areas: diagnosing the users, explaining pathogenesis and informing on therapy. The chatbot was freely available online³. Within 15 months, we collected 254 usable responses. The responses were split into varying groups for statistical analysis (Fig. 1).

For further details, see the supplementary material (for all on-line suppl. material, see www.karger.com/doi/10.1159/000511706) (Fig. 1) [7, 13, 14].

Results

Demographics

Over the period of 461 days, 285 calls on the website were recorded. Roughly 96% of inquirers entered the website directly through the IP address, not by link from a different website. Looking at devices, 65% used a smartphone and 31% a desktop computer.

In total, 254 cases were eligible for analysis, as they answered the item *interest*. Of those, 239 responses were of possible patients: 211 undiagnosed and 28 diagnosed. The group No_HS contained 15 users (Table 1).

The mean age of users (excluding No_HS and users who stated 99 years) was 32.49 years ($n = 229$) with a range from 16 to 80 years. Split into groups, 24.9% ($n = 57$) were below 25 years of age, 40.6% ($n = 93$) were between 25 and 35 years old and 34.5% ($n = 79$) were 35

² <https://trends.google.com/trends/explore/TIMESERIES/1582207200?hl=en-US&tz=-60&date=2010-01-20+2020-01-20&q=%2Fm%2F02rxy7&sn=3> (07.03.2020).

³ www.hs-bot.com (16.06.2020).

Table 1. Selected data points and analysed responses

	All users		Undiagnosed		Diagnosed		No_HS	
Interest	254	100%	211	83.07%	28	11.02%	15	5.91%
Mean age, years	32.7		32.4		33.3		36	
Age groups								
<25 years	59	24.58%	51	25.00%	6	24.00%	2	18.18%
25–34 years	97	40.42%	83	40.69%	10	40.00%	4	36.36%
≥35 years	84	35.00%	70	34.31%	9	36.00%	5	45.45%
Female	169/240	70.42%	139/202	68.81%	21/26	80.77%	9/12	75.00%
Number of jumps	14.25±8.83		14.66±8.55		14.86±10.90		7.33±5.31	
From main	12.33±7.94		12.59±7.62		12.48±10.48		6.78±3.93	
Tentative diagnosis	Affected		Not affected					
HS	149	88.69%	19	11.31%				
Crohn’s disease	16	13.56%	102	86.44%				
Acne vulgaris	45	45.00%	55	55.00%				
Dissecting cellulitis of the scalp	8	8.89%	82	91.11%				
Pilonidal sinus	17	21.25%	63	78.75%				
Pyoderma gangraenosum	2	2.63%	74	97.37%				
Hurley stage	I		II		III			
	16	39.02%	19	46.34%	6	14.63%		

HS, hidradenitis suppurativa. Noted are the percentages within the whole sample. Affected patients were identified through the diagnostic process and direct enquiry. Not affected are those for whom the illnesses could be ruled out.

years or older (Table 1). The 90th percentile was at 50 years.

Of the 239 respondents, 160 (70.2%) stated to be female, and 68 (29.8%) stated to be male (Fig. 2). No sex or gender was stated by 11 users.

User Interest

Overall, 209 undiagnosed and diagnosed users arrived at the item *main*. Of those, 165 (81.3%) chose *diagnostics*, 9 (4.4%) *pathogenesis*, 25 (12.3%) *therapy* and 4 (2.0%) *other_question* (Table 2). Amongst returning users, 1 (3.3%) chose *diagnostics*, 10 (33.3%) *pathogenesis*, 13 (43.3%) *therapy*, 3 (10%) *other_question* and 3 (10%) chose the added option of rating the chatbot.

Of 187 entries into the diagnostic branch, 19 cases did not report chronic lesions or had lesions in places atypical for HS. Another 19 responders could not be evaluated. Overall, 149 (79.7%) diagnoses were positive (Table 2). In total, 142 previously undiagnosed and an additional 28 diagnosed users added up to a minimum of 170 *alleged patients* among all 254 responders (66.9%).

Further along, 56 users underwent the staging process. Of those, 16 (39.0%) were diagnosed with Hurley stage I, 19 (46.3%) with stage II and 6 (14.6%) with stage III (Table 1). For 15 responders, staging was not possible.

Users typed 34 queries into the *other_question* text box. For 21 searches, the matching item showed up. For 13 enquiries no matching item was found, and users were led back to the item *main_oq*.

Number of Jumps

The average number of jumps (NOJ) from the start was 14.69 (±8.83). Blind jumps were not counted.

In the informational part, starting at the item *main*, the average NOJ was 12.57 (±7.99, $n = 207$). Undiagnosed users took 12.59 (±7.62) jumps, diagnosed users 12.48 (±10.48). No_HS users took 6.78 (±3.93, $n = 9$) jumps (Table 1). Among the first two groups, 43 (14.6%) users took 4 or 5 jumps from *main*.

Alleged patients took more jumps (mean = 13.82, SD = 8.02, $n = 167$) than unaffected users (mean = 9.11,

Table 2. Analysis of the jumps from main by stated interest, age groups and sex

Main, by interest	Undiagnosed		Diagnosed		No_HS	
Diagnostics	164	92.13%	1	4.00%	3	33.33%
Pathogenesis	3	1.69%	6	24.00%	0	0.00%
Therapy	9	5.06%	16	64.00%	3	33.33%
Other question	2	1.12%	2	8.00%	3	33.33%
Main, by age group	<25 years		25–34 years		≥35 years	
Diagnostics	38	77.55%	74	85.06%	51	80.95%
Pathogenesis	3	6.12%	4	4.60%	1	1.59%
Therapy	6	12.24%	8	9.20%	11	17.46%
Other question	2	4.08%	1	1.15%	0	0.00%
Positive tentative diagnosis	41 of 44	74 of 82	51 of 57			
Main, by sex	Female		Male			
Diagnostics	108	80.60%	54	83.08%		
Pathogenesis	6	4.48%	3	4.62%		
Therapy	16	11.94%	8	12.31%		
Other question	4	2.99%	0	0.00%		
Jumps from main	12.82±8.05	12.58±7.88

Fig. 2. Flowchart with the first few items (italics) and number of responses. The boxes to the right contain the particular options on the item. The paths split up after main. From there, jumps could also lead back to previous items. NOJ, number of jumps.

SD = 5.37, $n = 27$), $t(47.1) = 3.91$, $p < 0.01$. The effect size along Cohen is $r = 0.50$ and thus shows a strong effect.

Discussion/Conclusion

There are over 95 million native German speakers in the world [15]. Internet user penetration in Europe is 80.6% [16]. With a rough prevalence of 1% for HS, there are over 760,000 people in the target group of the chatbot. In addition, an even larger group will benefit from being informed about HS and excluding the diagnosis for themselves.

HS typically develops during adolescence. It is a chronic disease, but the activity usually decreases after 10–15 years. Nonetheless, the chatbot users were distributed around a younger age than the complete patient collective (Table 1). This reflects the dispersion of use of digital health and the Internet in general. Moreover, compared to data collected by the Swiss federal statistical office [17], the users of the HS Chatbot were on average even younger than general consumer health informatics users.

The age of users has been seen to guide their interest: people under 25 years old were more interested in pathogenesis (7 jumps from main, $n = 57$) than the rest (11 jumps, $n = 172$) (12.3 vs. 6.4%). People over 35 were more interested in therapy (17 jumps, $n = 79$; rest: 21 jumps, $n = 160$; 21.5 vs. 13.1%).

No other elements (Hurley stage, NOJ, etc.) were influenced by age or sex.

The large majority of users belonged to the undiagnosed group. Out of those, 92.13% (164/178) chose diagnostics at the item *main* (Table 2). Clearly, offering a diagnosis was the most important function of the chatbot.

Of those who completed the diagnostic track, 84.1% (138/164) received a positive tentative diagnosis. As the disease is not commonly known, people with uncharacteristic lesions do not arrive at the platform. Most users probably suspect being affected by HS and might be drawn to answers that confirm their suspicion (confirmation bias).

Stage II was slightly more common among chatbot users than stage I, while stage III was far less common (Table 1). In reality, stage I is more frequent than stage II [10]. Firstly, the staging mechanism may be inaccurate. Secondly, there is a bias of exaggeration when classifying one's own lesions. Thirdly, there is a bigger strain for people with a later stage.

Stage III was more common in male users ($p < 0.05$ in Fisher's exact test, $\phi = 0.36$ [$p < 0.05$]), which is in line with the findings of Schrader et al. [18].

Potential patients made almost twice as many jumps as unaffected users (Table 1). So did the alleged patients compared to the rest of the undiagnosed users. People affected by HS have a stronger interest in the disease than healthy people.

The most frequented path led to the item *diagnosis_true* by 4 jumps from *main*. It was taken by 71.5% (138/193) of the users who took sufficient jumps. This is the point of a tentative diagnosis. It corresponds to the peak in the NOJ at 4–5 jumps. The replies on later items drop quickly: 18.1% (25/138) quit either at *diagnosis_true* or at one of the subsequent items. This shows that users are likely to leave the website when the desired answer is found.

On average, each item received 41.6 responses (median = 11).

At the item *rating*, one user mentioned that some items were too complex, while others wished for more information. A potential bias is that only committed users filled in the item.

General Applications

Essential questions for clinical use of an online platform concern main interests, subgroups of common interest, presumed NOJ, demographics, access and health informatics skills.

It was striking how many people used the chatbot to get a presumptive diagnosis. Therefore, diagnostics may be hierarchically put above pathogenesis and therapy. They have to detect all possibly affected users within a small number of clicks and refer them to consultation by a doctor. The process has to be accurate but not drawn out. To be more reliable, the methods used need to be improved, or an entirely different approach may be necessary.

All users should be asked about previously run diagnostics, disease staging and other diagnoses in the beginning. To analyse the accuracy of the chatbot, these specifications would have to be known.

The average NOJ has to be considered. The fundamental contents must be incorporated before a predicted peak in the NOJ (in this case, stating a diagnosis within 5 jumps). This is when users have to be engaged to further commit to the chatbot. The users of the HS Chatbot who started the diagnostic process mostly answered all necessary items. Generally, if users are working towards a specific goal, they rarely quit. Progress bars along the indi-

vidual branches may make users stick to a process until the end. Smaller text boxes are less overwhelming and feel less cluttered. Specific instructions and a tutorial in the beginning may help with the understanding. Furthermore, it should always be stated what specific personal information is used for.

When working with predefined jumps, as in this project, a prediction has to be made on what users want to know, following each item. Matching options have to be offered. Items could be combined in response to certain requests. Particularly, all associated diseases could be summed up together and a general referral item could be added where necessary. Users should be able to revise past items and select different options.

Conclusions

In prospect, platforms based on artificial intelligence and machine learning are promising options to complement medicine and extend physicians' reach to as yet undiagnosed patients. In future versions of more advanced chatbots, it would be useful if machine learning algorithms could understand and react to free text responses. Jumps could be adaptive and self-enhancing regarding diagnosis. Once set up, they would be easily applicable to any illness without requiring deep technological insight by the creator submitting content. An unrestricted system is more applicable to clinical use, as the individually generated responses are difficult to correlate. For more representative scientific data, the chatbot should be tested in a population of selected patients.

Key Message

Chatbots, interactive conversational platforms based on artificial intelligence, can provide specific information and simultaneously facilitate patient empowerment and management.

Acknowledgement

Sarah Piccirillo and other members of the Verein Acne Inversa SchwAIz generously contributed the patients' view.

The chatbot was distributed through a link on the homepage of AI SchwAIz⁴ and by flyers and presentations in the following hospitals and doctors' offices: Haut. Venen. Allergie. Zentrum Brunehof, Uster; Luzerner Kantonsspital, Luzern; Spital Thurgau AG, Frauenfeld; Universitätsspital Zurich, Zurich.

The homepage and CMS were provided by Swiss4ward.

Helen Mawer of the Department of Dermatology Basel organised the submission of the paper.

Statement of Ethics

Ethical approval was sought with Kantonale Ethikkommission Zurich, and confirmation of freedom to operate with anonymous data was received (decision No. BASEC-2018-00937). A disclaimer and consent form on the welcome page informed about the data that were collected for analysis. There was no contact between the authors and the users.

Conflict of Interest Statement

Prof. Navarini has received consulting fees from various companies producing drugs for HS. Vahid Djamei is the proprietor of Swiss4ward, the company developing the online questionnaires used in this study.

Funding Sources

There were no funding sources.

Author Contributions

The contents of the HS Chatbot were chosen and phrased by M. Walss. He analysed the derived data and composed this paper. Prof. N. Navarini initiated the study by identifying the public need for information on HS. He provided knowledge on both the disease and digital health resources. Initial medical proficiency and the photographs for the chatbot were kindly contributed by Dr. F. Anzengruber. V. Djamei made valuable contributions on the design of the website and data extraction. Dr. A. Arafa helped by advising on and conducting some statistical analysis.

References

- 1 Demir G. Consumer health informatics: past, present, and future of a rapidly evolving domain. *Yearb Med Informat.* 2016;Suppl 1(Suppl 1):S42–7.
- 2 Arafa AE, Anzengruber F, Mostafa AM, Navarini AA. Perspectives of online surveys in dermatology. *J Eur Acad Dermatol Venereol.* 2019 Mar;33(3):511–20.
- 3 Coulter A, Ellins J, Swain D, Clarke A, Heron P, Rasul F, et al. Assessing the quality of information to support people in making decisions about their health and healthcare. Oxford: Picker Institute Europe; 2006.
- 4 Petrič G, Atanasova S, Kamin T. Ill Literates or Illiterates? Investigating the eHealth Literacy of Users of Online Health Communities. *J Med Internet Res.* 2017 Oct;19(10):e331.

⁴ <https://acneinversaschweiz.ch/ueber-acne-inversa/>.

- 5 Hill J, Ford WR, Farreras IG. Real conversations with artificial intelligence. *Comput Human Behav*. 2015;49(C):245–50.
- 6 Chalela P, Munoz E, Inupakutika D, Kaghyan S, Akopian D, Kaklamani V, et al. Corrigendum to “Improving adherence to endocrine hormonal therapy among breast cancer patients: study protocol for a randomized controlled trial” [Contemp Clin Trials Commun. 2018;12:109–115]. *Contemp Clin Trials Commun*. 2019 Mar;14:100349.
- 7 Jemec GB. *Hidradenitis suppurativa*. Berlin: Springer; 2006.
- 8 Saunte DM, Jemec GB. Hidradenitis Suppurativa: Advances in Diagnosis and Treatment. *JAMA*. 2017 Nov;318(20):2019–32.
- 9 Hurley H. *Dermatologic surgery, principles and practice*. New York: Dekker; 1989. p. 729–39.
- 10 Canoui-Poitine F, Revuz JE, Wolkenstein P, Viallette C, Gabison G, Pouget F, et al. Clinical characteristics of a series of 302 French patients with hidradenitis suppurativa, with an analysis of factors associated with disease severity. *J Am Acad Dermatol*. 2009 Jul;61(1): 51–7.
- 11 Saunte DM, Boer J, Stratigos A, Szepletowski JC, Hamzavi I, Kim KH, et al. Diagnostic delay in hidradenitis suppurativa is a global problem. *Br J Dermatol*. 2015 Dec;173(6): 1546–9.
- 12 Hessam S, Salem J, Bechara FG, Haferkamp A, Heidenreich A, Paffenholz P, et al. Hidradenitis suppurativa gains increasing interest on World Wide Web: a source for patient information? *Int J Dermatol*. 2017 Jul;56(7):726–32.
- 13 Eid M, Schmidt K. *Testtheorie und Testkonstruktion*. Goettingen: Hogrefe Verlag; 2014.
- 14 Israel GD, Taylor CL. Can response order bias evaluations? *Eval Program Plann*. 1990;13(4): 365–71.
- 15 Ammon U. *Die Stellung der deutschen Sprache in der Welt*. Berlin: De Gruyter; 2014.
- 16 The World Bank. World development indicators: the information society [cited 2020 Mar 8]. Available from: <http://wdi.worldbank.org/table/5.12#>
- 17 Bundesamt für Statistik BFS. STAT-TAB – interaktive Tabellen: Einzelpersonen und Online-Aktivitäten [cited 2020 Mar 8]. Available from: https://www.pxweb.bfs.admin.ch/pxweb/de/px-x-1604000000_102/px-x-1604000000_102/px-x-1604000000_102.px/table/tableViewLayout2/?rxid=de5d516a-56dd-4951-b3c7-5ffd36607476
- 18 Schrader AM, Deckers IE, van der Zee HH, Boer J, Prens EP. Hidradenitis suppurativa: a retrospective study of 846 Dutch patients to identify factors associated with disease severity. *J Am Acad Dermatol*. 2014 Sep;71(3): 460–7.