RESEARCH LETTER

Online symptom checkers lack diagnostic accuracy for skin rashes

To the Editor: Online symptom checkers allow individuals to enter signs and symptoms to self-diagnose illnesses. Prior studies outside of dermatology show that online symptom checkers have poor diagnostic accuracy.[1-4] We performed a cross-sectional comparative study of 8 online symptom checkers' diagnostic accuracies for skin rashes and determined the minimum number of user-input data terms required for correct diagnoses.

Fifteen rash vignettes were independently reviewed by 3 board-certified dermatologists. The diagnoses included atopic dermatitis (pediatric), cellulitis, dermatitis herpetiformis, erythema nodosum, folliculitis, hand-foot-mouth disease (pediatric), herpes simplex (genital), hidradenitis suppurativa, lichen sclerosus, perioral dermatitis, pityriasis rosea, psoriasis, rosacea, seborrheic dermatitis, and shingles. We entered vignette clinical data terms into 8 online symptom checkers. Six were from a recent review article in 2015.[2] Two were newer symptom checkers most reviewed on the Apple App Store. Symptom checkers needed to still exist (since 2015 study), be capable of diagnosis (not just triage), allow user input of skin rash data, and handle pediatric and adult ages.

Since symptom checkers provide multiple potential diagnoses, we scored a correct diagnosis when the vignette diagnosis was the first (primary) diagnosis of the symptom checker. When correctly diagnosed, we tested the symptom checker fidelity by randomly removing clinical terms one by one, until the symptom checker produced an incorrect diagnosis, and determined the minimum number of data terms required. We performed 10 iterations of

this process for each correctly diagnosed rash to enhance randomization and simulate various patients entering incomplete data.

Online symptom checkers correctly diagnosed skin rashes in only 37 of 120 entries (30.8%; Table I), with accuracy in the range of prior studies from other fields (26% to 36%).[1-4] In those 37 scenarios, symptom checkers generated a mean total of 13.6 differential diagnoses (SD 16.3, range: 1-72). When symptom checkers were correct, the mean number of data entry terms needed to maintain that correct diagnosis, after one by one random elimination, was 6.1 (SD: 2.4, range: 1-16). There was variation in each symptom checker's accuracies, ranging from 13.3% to 73.3%. The tested symptom checkers correctly diagnosed 75% of atopic dermatitis (pediatric) and 62.5% of shingles (Table II). No symptom checker correctly diagnosed dermatitis herpetiformis, and only one symptom checker correctly identified erythema nodosum, hidradenitis suppurativa, lichen sclerosus, or perioral dermatitis.

Accuracy differences may reflect algorithms' broad scopes of symptoms, signs, and organ systems (not dermatology-specific) and varying queries for pertinent history such as gluten sensitivity and medications. This study is limited in that other dermatologic conditions may lead to different diagnostic outcomes. We chose our 15 vignettes, verified by 3 board-certified dermatologists, attempting to maximize the opportunities for correct diagnosis.

For diagnosing skin rashes, symptom checkers have poor accuracy due to limitations on asking about pertinent medical history and insufficient diagnoses in their algorithms. Even when correct, symptom checkers' broad differentials and low specificity added significant diagnostic uncertainty,

Table I. Performance of each online symptom checker in diagnosing skin rashes

Symptom checker	Correct diagnosis generated by symptom checker (%)	Correct diagnosis within top 3 diagnoses generated by symptom checker (%)	Correct diagnosis within symptom checkers' generated differential (%)	Correct diagnosis generated by symptom checker when limited to generated differential diagnoses (%)
AskMD/Sharecare	2/15 (13.3)	7/15 (46.7)	10/15 (66.7)	2/10 (20)
Esagil	3/15 (20)	3/15 (20)	3/15 (20)	3/3 (100)
Isabel	3/15 (20)	5/15 (33.3)	12/15 (80)	3/12 (25)
Symptify	2/15 (13.3)	2/15 (13.3)	2/15 (13.3)	2/2 (100)
Symptomate	3/15 (20)	5/15 (33.3)	5/15 (33.3)	3/5 (60)
WebMD	8/15 (53.3)	10/15 (66.7)	14/15 (93.3)	8/14 (57.1)
Ada*	11/15 (73.3)	14/15 (93.3)	14/15 (93.3)	11/14 (78.6)
K Health*	5/15 (33.3)	6/15 (40)	6/15 (40)	5/6 (83.3)
All	37/120 (30.8)	52/120 (43.3)	66/120 (55)	37/66 (56.1)

*Apple App Store mobile applications.

Table II. Online symptom checker diagnostic accuracy for each rash

Diagnosis	Correct diagnosis (%)
Atopic dermatitis*	6/8 (75)
Shingles	5/8 (62.5)
Psoriasis	4/8 (50)
Rosacea	4/8 (50)
Hand-foot-mouth disease*	3/8 (37.5)
Herpes simplex (genital)	3/8 (37.5)
Cellulitis	2/8 (25)
Folliculitis	2/8 (25)
Pityriasis rosea	2/8 (25)
Seborrheic dermatitis	2/8 (25)
Erythema nodosum	1/8 (12.5)
Hidradenitis suppurativa	1/8 (12.5)
Lichen sclerosus	1/8 (12.5)
Perioral dermatitis	1/8 (12.5)
Dermatitis herpetiformis	0/8 (0)

*Pediatric cases.

despite the input of multiple data parameters. With widespread use of online health information, it is valuable to highlight that symptom checkers are not currently useful for skin rash self-diagnoses.

Nicholas A. Berry, BS,[a] Jamison A. Harvey, MD,[b] Mark R. Pittelkow, MD,[b] David L. Swanson, MD,[b] and Yul W. Yang, MD, PhD[b]

From the Mayo Clinic Alix School of Medicine Arizona, Scottsdale, Arizona[a]; Department of Dermatology at Mayo Clinic Arizona, Scottsdale, Arizona.[b]

Conflicts of interest

None disclosed.

REFERENCES
1. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. *Med J Aust*. 2020;212(11):514-519. https://doi.org/10.5694/mja2.50600
2. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480. https://doi.org/10.1136/bmj.h3480
3. Shen C, Nguyen M, Gregor A, et al. Accuracy of a popular online symptom checker for ophthalmic diagnoses. *JAMA Ophthalmol*. 2019;137(6):690-692. https://doi.org/10.1001/jamaophthalmol.2019.0571
4. Berry AC, Berry NA, Wang B, et al. Symptom checkers versus doctors: a prospective, head-to-head comparison for cough. *Clin Respir J*. 2020;14(4):413-415. https://doi.org/10.1111/crj.13135