Original Investigation

Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma

A Web-Based International Dermoscopy Society Study

Cristina Carrera, MD, PhD; Michael A. Marchetti, MD; Stephen W. Dusza, DrPH; Giuseppe Argenziano, MD; Ralph P. Braun, MD; Allan C. Halpern, MD; Natalia Jaimes, MD; Harald J. Kittler, MD; Josep Malvehy, MD; Scott W. Menzies, MBBS, PhD; Giovanni Pellacani, MD; Susana Puig, MD; Harold S. Rabinovitz, MD; Alon Scope, MD; H. Peter Soyer, MD; Wilhelm Stolz, MD; Rainer Hofmann-Wellenhof, MD; Iris Zalaudek, MD; Ashfaq A. Marghoob, MD

IMPORTANCE The comparative diagnostic performance of dermoscopic algorithms and their individual criteria are not well studied.

OBJECTIVES To analyze the discriminatory power and reliability of dermoscopic criteria used in melanoma detection and compare the diagnostic accuracy of existing algorithms.

DESIGN, SETTING, AND PARTICIPANTS This was a retrospective, observational study of 477 lesions (119 melanomas [24.9%] and 358 nevi [75.1%]), which were divided into 12 image sets that consisted of 39 or 40 images per set. A link on the International Dermoscopy Society website from January 1, 2011, through December 31, 2011, directed participants to the study website. Data analysis was performed from June 1, 2013, through May 31, 2015. Participants included physicians, residents, and medical students, and there were no specialty-type or experience-level restrictions. Participants were randomly assigned to evaluate 1 of the 12 image sets.

MAIN OUTCOMES AND MEASURES Associations with melanoma and intraclass correlation coefficients (ICCs) were evaluated for the presence of dermoscopic criteria. Diagnostic accuracy measures were estimated for the following algorithms: the ABCD rule, the Menzies method, the 7-point checklist, the 3-point checklist, chaos and clues, and CASH (color, architecture, symmetry, and homogeneity).

RESULTS A total of 240 participants registered, and 103 (42.9%) evaluated all images. The 110 participants (45.8%) who evaluated fewer than 20 lesions were excluded, resulting in data from 130 participants (54.2%), 121 (93.1%) of whom were regular dermoscopy users. Criteria associated with melanoma included marked architectural disorder (odds ratio [OR], 6.6; 95% CI, 5.6-7.8), pattern asymmetry (OR, 4.9; 95% CI, 4.1-5.8), nonorganized pattern (OR, 3.3; 95% CI, 2.9-3.7), border score of 6 (OR, 3.3; 95% CI, 2.5-4.3), and contour asymmetry (OR, 3.2; 95% CI, 2.7-3.7) ($P < .001$ for all). Most dermoscopic criteria had poor to fair interobserver agreement. Criteria that reached moderate levels of agreement included comma vessels (ICC, 0.44; 95% CI, 0.40-0.49), absence of vessels (ICC, 0.46; 95% CI, 0.42-0.51), dark brown color (ICC, 0.40; 95% CI, 0.35-0.44), and architectural disorder (ICC, 0.43; 95% CI, 0.39-0.48). The Menzies method had the highest sensitivity for melanoma diagnosis (95.1%) but the lowest specificity (24.8%) compared with any other method ($P < .001$). The ABCD rule had the highest specificity (59.4%). All methods had similar areas under the receiver operating characteristic curves.

CONCLUSIONS AND RELEVANCE Important dermoscopic criteria for melanoma recognition were revalidated by participants with varied experience. Six algorithms tested had similar but modest levels of diagnostic accuracy, and the interobserver agreement of most individual criteria was poor.

JAMA Dermatol. 2016;152(7):798-806. doi:10.1001/jamadermatol.2016.0624
Published online April 13, 2016.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Ashfaq A. Marghoob, MD, Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, 16 E 60th St, New York, NY 10022 (marghooa@mskcc.org).

Use of dermoscopy by trained users, but not novices, improves diagnostic accuracy for cutaneous melanoma compared with naked eye examination alone.[1] Experts of dermoscopy tend to review a dermoscopic image and reach a diagnosis without use of structured analytical criteria, a diagnostic process that can be referred to as pattern analysis. Multiple simplified dermoscopic algorithms, such as the ABCD rule, the Menzies method, the 7-point checklist, the 3-point checklist, chaos and clues, and CASH (color, architecture, symmetry, and homogeneity), were developed to facilitate a novice's ability to distinguish melanomas from nevi with high diagnostic accuracy.[2-7] A comparison of these algorithms reveals 2 diverging approaches to simplified melanoma detection (**Table 1**). The ABCD rule and CASH principally quantify the overall organization of a lesion by assessing features such as symmetry, architectural disorder, border sharpness, and heterogeneity in colors and structures. However, the 7-point checklist relies on the identification of atypical appearances of dermoscopic structures (eg, atypical network) in distinction from otherwise normal counterparts or on identifying unique structures strongly associated with melanoma (eg, regression). Chaos and clues, the Menzies method, and the 3-point checklist include elements of both approaches.

Although each algorithm has unique criteria, there is significant overlap in their concepts, which may explain why the ABCD rule, the Menzies method, and the 7-point checklist have similar overall accuracy in the diagnosis of melanocytic lesions by novices.[8] Beginners and instructors of dermoscopy are consequently unclear as to which, if any, algorithm(s) they should use and teach, respectively. In addition, no algorithm has been significantly revised since its initial publication to include newly identified dermoscopic features with high specificity for melanoma, such as negative network or white shiny structures.[9,10] A critical need exists to better understand the comparative diagnostic performance of dermoscopic algorithms, in particular the discriminatory power and interobserver agreement of their individual criteria. The primary objective of this study was to measure the discriminatory power and interobserver agreement of individual dermoscopic criteria, including newly described dermoscopic features. A secondary objective was to compare the diagnostic accuracy of 6 existing simplified algorithms.

Key Points

Question What is the discriminatory power and reliability of dermoscopic criteria used in melanoma detection?

Findings In this survey-based study, the diagnostic importance of new and previously identified dermoscopic criteria for melanoma detection was validated; however, the majority of criteria had poor to fair interobserver agreement. Criteria with relatively strong discriminatory power and moderate levels of interobserver agreement included architectural disorder, pattern asymmetry, contour asymmetry, comma vessels, and absence of vessels.

Meaning Further efforts are needed to standardize terminology and definitions of dermoscopic criteria.

Methods

The Memorial Sloan Kettering Cancer Center Institutional Review Board approved this study without the requirement for written informed consent in accordance with the Helsinki Declaration. Data were deidentified.

Lesion Selection

Twelve pigmented lesion clinics from Australia, Austria, Germany, Italy, Spain, Switzerland, and the United States contributed study images. Each contributor provided up to 50 lesions with a 1:3 ratio of melanomas to nevi. Melanomas were required to have an unequivocal histopathologic diagnosis, and

Table 1. Comparison of Dermoscopic Criteria of Simplified Diagnostic Algorithms for Melanoma

Criterion	ABCD Rule	CASH	Menzies Method	7-Point Checklist	3-Point Checklist	Chaos and Clues
Symmetry in colors or structures	✓	✓	✓		✓	✓
Border sharpness	✓					
Quantity of specified colors	✓	✓	✓			
Quantity of specified structures	✓[a]	✓[b]				
Architectural disorder		✓				
Blue-white veil			✓	✓	✓	✓
Any blue or white color					✓	✓
Atypical dots or globules			✓	✓		✓
Regression			✓	✓	✓	✓
Streaks			✓	✓		✓
Atypical network			✓	✓	✓	✓
Atypical vessels				✓		✓
Irregular blotch				✓		✓

Abbreviation: CASH, color, architecture, symmetry, and homogeneity.

[a] The ABCD rule includes dots, globules, structureless areas, network, and streaks and does not distinguish between atypical and typical structures.

[b] CASH includes dots or globules, blotches, network, regression, streaks, blue-white veil, and polymorphous vessels and does not distinguish between atypical and typical structures.

nevi were required to be histopathologically verified or to have demonstrated stability under sequential dermoscopic imaging over time. Contributors sequentially selected lesions from their patient records and used 1:1 randomization of lesions into polarized vs nonpolarized sets. Other requested data included anatomical location, patient age and sex, imaging modality (polarized vs nonpolarized), and a clinical close-up image.

A total of 580 lesions (140 melanomas and 440 nevi) were contributed to the study. Lesions were reviewed by Memorial Sloan Kettering Cancer Center investigators, and 103 were excluded because of (1) location on acral, mucosal, or facial sites, (2) inadequate image quality, (3) equivocal diagnosis after review of the pathology report or sequential imaging, (4) nonmelanocytic lesions, and (5) lesions from patients younger than 18 years. The final data set was composed of 477 unique lesions, of which 119 (24.9%) were melanomas. Lesions were randomized into 12 image sets that contained 39 (n = 8) or 40 (n = 7) unique lesions and 5 nonunique lesion images (2 melanoma, 3 benign) that were repeated in all sets.

Web-Based Study Interface

Algorithm tutorials were created and posted by dermoscopic experts through the International Dermoscopy Society (IDS) website. Review of tutorials was encouraged but not mandatory for participants, and links to tutorials were available on the main study site interface and the data collection form.

Participant Selection

A link present on the IDS website from January 1, 2011, through December 31, 2011, directed participants to the study website (www.dermoscopy-ids.org). Data analysis was performed from June 1, 2013, through May 31, 2015. Participation was open to attending physicians, residents, and medical students and was not restricted by specialty type or experience level. Image contributors were excluded from the study. Participants were required to register and specify their specialty, years of clinical experience, preferred dermoscopic analysis method, dermoscopy frequency of use, predominant modality (polarized vs nonpolarized) of use, and experience. There was no incentive for study participation.

Two hundred forty participants registered for the study, and 103 (42.9%) completed all available images in their data sets. The 110 participants (45.8%) who evaluated fewer than 20 lesions were excluded, resulting in data from a total of 130 participants (54.2%) eligible for analysis.

Participant Evaluation

A comprehensive list of all dermoscopic structures from the dermoscopy algorithms was created, and overlapping criteria were merged into 1 criterion (eg, granularity and peppering were combined into 1 criterion). Newly identified dermoscopic structures with high specificity for melanoma (eg, negative network, chrysalis structures [shiny white or crystalline structures], polymorphous vessels, atypical vessels, and pink veil) were included. Criteria included (1) global pattern, (2) pattern organization, (3) symmetry of contour, (4) symmetry of pattern, (5) architectural disorder, (6) abruptness of lesion border, (7) colors, and (8) melanocytic structures, including network and vascular structures. Participants examined the close-up clinical image of each lesion before viewing the dermoscopic image. The modality (polarized vs nonpolarized) of dermoscopic images was specified. There were no time constraints. For each lesion, the participant indicated the presence or absence of all dermoscopic criteria on the same webpage. Users were unable to modify their responses for a lesion after submission of data.

Statistical Analysis

Descriptive statistics and graphic methods were used to describe participant and lesion characteristics and participant dermoscopic evaluations because block randomization was used and no participants evaluated all images. Data were assessed as individual dermoscopic evaluations and as consensus evaluations for participants who reviewed a given study lesion. For individual evaluations, prevalence of each dermoscopic feature was tabulated along with 95% CIs. To quantify the association for the presence or absence of each feature with melanoma status, tabular cross-classifications, χ^2 statistics, and the associated odds ratios (ORs) and 95% CIs were calculated. Robust SEs were estimated to adjust for the clustered observations within reviewers. Intraclass correlation coefficients (ICCs) were estimated for each dermoscopic feature using 2-way random-effects models, with the dermoscopic raters treated as a random effect. This approach assumes that raters are randomly sampled from the larger population of raters with dermoscopic experience. The ICC is equal to 0 when the agreement is exactly what is expected by chance and 1 when there is perfect agreement. Intermediate values were interpreted as follows: poor, 0.01 to less than 0.2; fair, 0.2 to less than 0.4; moderate, 0.4 to less than 0.6; substantial, 0.6 to less than 0.8; and almost perfect agreement, greater than 0.8.

For consensus evaluations, the presence or absence of each dermoscopic feature was calculated as the proportion of participants who identified the feature for a given lesion. When 50% or more of the participants identified a dermoscopic feature for a given study lesion, the attribute was considered present. We applied consensus evaluations to dermoscopic algorithms to evaluate performance. Using logistic regression models with the dichotomous outcome of melanoma vs nevus, we compared areas under the receiver operating characteristic (ROC) curve among the diagnostic algorithms. Analyses were performed with STATA statistical software, version 12.1 (StataCorp).

Results

Participants

The 130 participants who evaluated 20 lesions or more had a mean (SD) of 12 (8.7) years of dermatology experience. The mean (SD) percentages of their practice that was composed of skin cancer screening and the population at high risk for skin cancer were 33.5% (25.8%) and 14.4% (16.4%), respectively. A total of 73 participants (56.2%) reported being attending

dermatologists, 122 (93.8%) were comfortable using dermoscopy, and 121 (93.1%) were regular users of dermoscopy (**Table 2**).

Lesion Evaluations

A total of 477 unique lesions were evaluated in the study. Each lesion was evaluated by a median of 12 participants, with the exception of the 5 lesions that were repeated in the 12 image sets and evaluated by all 130 participants, resulting in a total of 5670 unique lesion evaluations.

Interobserver Agreement of Dermoscopic Criteria

Most dermoscopic criteria had poor to fair interobserver agreement, including features such as atypical network (ICC, 0.21; 95% CI, 0.17-0.25), blue-white veil (ICC, 0.34; 95% CI, 0.30-0.39), regression (ICC, 0.11; 95% CI, 0.08-0.13), and atypical vessels (ICC, 0.26; 95% CI, 0.22-0.30) (**Table 3**).

Criteria with moderate levels of interobserver agreement included comma vessels (ICC, 0.44; 95% CI, 0.40-0.49), absence of vessels (ICC, 0.46; 95% CI, 0.42-0.51), dark brown color (ICC, 0.40; 95% CI, 0.35-0.44), and architectural disorder (ICC, 0.43; 95% CI, 0.39-0.48) (Table 3). Absence of network (ICC, 0.39; 95% CI, 0.34-0.43), pattern symmetry (ICC, 0.37; 95% CI, 0.32-0.41), contour symmetry (ICC, 0.37; 95% CI, 0.32-0.42), and total colors present (ICC, 0.36; 95% CI, 0.31-0.40) had similar levels of interobserver agreement.

Dermoscopic Criteria Associated With Melanoma Status

Criteria strongly associated with melanoma status (OR ≥3) included marked architectural disorder (OR, 6.6; 95% CI, 5.6-7.8), pattern asymmetry (OR, 4.9; 95% CI, 4.1-5.8), nonorganized pattern (OR, 3.3; 95% CI, 2.9-3.7), border score of 6 (OR, 3.3; 95% CI, 2.5-4.3), contour asymmetry (OR, 3.2; 95% CI, 2.7-3.7), polymorphous vessels (OR, 3.1; 95% CI, 2.4-4.0), border score of 5 (OR, 3.1; 95% CI, 2.3-4.2), and atypical vessels (OR, 3.0; 95% CI, 2.5-3.6) ($P < .001$ for all) (Table 3). Inability to determine features such as border score (OR, 4.1; 95% CI, 3.1-5.4), pattern symmetry (OR, 6.3; 95% CI, 3.6-10.8), and contour symmetry (OR, 6.3; 95% CI, 4.0-9.9) were also strongly associated with melanoma status (all $P < .001$). Other criteria associated with melanoma status are given in Table 3.

Criteria with a strong inverse association with melanoma status (OR <0.7) included comma vessels (OR, 0.4; 95% CI, 0.3-0.6), peripheral reticular with central hyperpigmentation global pattern (OR, 0.5; 95% CI, 0.4-0.6), globular global pattern (OR, 0.5; 95% CI, 0.4-0.6), 2-component symmetric global pattern (OR, 0.5; 95% CI, 0.3-0.7), regular brown dots (OR, 0.5; 95% CI, 0.4-0.6), regular brown globules (OR, 0.5; 95% CI, 0.4-0.7), absence of vessels (OR, 0.5; 95% CI, 0.4-0.5), regular blotch (OR, 0.4; 95% CI, 0.3-0.6), and light brown color (OR, 0.6; 95% CI, 0.5-0.7) (all $P < .001$) (Table 3).

The dermoscopic criteria with ICC levels of 0.37 or higher and relatively strong discriminatory power (OR ≥3.0 or <0.7) included comma vessels, absence of vessels, marked architectural disorder, pattern asymmetry, and contour asymmetry.

Table 2. Participant Characteristics

Characteristic	No. (%) (n = 130)
Clinical specialty	
Dermatologist	73 (56.2)
General practitioner	24 (18.5)
Dermatology resident	25 (19.2)
Medical student	1 (0.8)
Other	7 (5.4)
Do you regularly use dermoscopy?	
No	9 (6.9)
Yes	121 (93.1)
Dermoscopy modality used?	
Nonpolarized	41 (31.5)
Polarized	89 (68.5)
Comfortable practicing without dermoscopy?	
No	111 (85.4)
Yes	19 (14.6)
Comfortable using dermoscopy?	
No	8 (6.2)
Yes	122 (93.8)
Frequency of dermoscopy use?	
Almost always	118 (90.8)
Sometimes	5 (3.8)
Rarely	7 (5.4)
What do you use dermoscopy on?	
Most lesions	76 (58.5)
Selected lesions	17 (13.1)
Selected lesion plus few more	37 (28.5)
Preferred dermoscopy method?	
Pattern analysis	65 (50.0)
ABCD rule	19 (14.6)
7-Point checklist	13 (10.0)
3-Point checklist	10 (7.7)
Menzies method	9 (6.9)
Chaos and clues	6 (4.6)
CASH algorithm	2 (1.5)
Nonselective screening	1 (0.8)
Overall gestalt based on familiarity	1 (0.8)
7-Point checklist and pattern analysis	1 (0.8)
ABCD rule and pattern analysis	1 (0.8)
Do not own a dermoscope	1 (0.8)
No response	1 (0.8)
Do you use photography to follow up patients?	
No	22 (16.9)
Yes	108 (83.1)

Abbreviation: CASH, color, architecture, symmetry, and homogeneity.

Newly Identified Dermoscopic Criteria

Negative network (OR, 1.4; 95% CI, 1.1-1.8; $P = .005$) and white shiny structures (OR, 2.5; 95% CI, 1.8-3.5; $P < .001$) were significantly associated with melanoma status. However, both had poor interobserver agreement levels (negative network: ICC, 0.15; 95%, CI 0.12-0.18; white shiny structures: ICC, 0.16; 95% CI, 0.13-0.19).

Table 3. Association Between Dermoscopic Criteria With Melanoma Status

Dermoscopic Criterion	No. (%) of Lesions		OR (95% CI)	*P* Value	ICC (95% CI)[a]
	Nevus (n = 4064)	Melanoma (n = 1541)			
Global pattern					
Diffuse reticular: present	720 (17.7)	215 (14.0)	0.8 (0.6-0.9)	.001	0.25 (0.21-0.29)
Patchy reticular: present	481 (11.8)	173 (11.2)	0.9 (0.8-1.1)	.53	0.17 (0.14-0.20)
Peripheral reticular with central hypopigmentation: present	306 (7.5)	108 (7.0)	0.9 (0.7-1.2)	.50	0.32 (0.28-0.37)
Peripheral reticular with central hyperpigmentation: present	481 (11.8)	97 (6.3)	0.5 (0.4-0.6)	<.001	0.29 (0.24-0.33)
Peripheral reticular with central globules: present	159 (3.9)	41 (2.7)	0.7 (0.5-1)	.02	0.13 (0.10-0.16)
Homogeneous: present	324 (8.0)	126 (8.2)	1.0 (0.8-1.3)	.80	0.22 (0.18-0.25)
Peripheral globular: present	168 (4.1)	43 (2.8)	0.7 (0.5-0.9)	.02	0.32 (0.28-0.36)
Globular: present	317 (7.8)	60 (3.9)	0.5 (0.4-0.6)	<.001	0.28 (0.24-0.32)
Multicomponent: present	157 (3.9)	75 (4.9)	1.3 (1.0-1.7)	.09	0.05 (0.03-0.06)
Two-component symmetric: present	166 (4.1)	32 (2.1)	0.5 (0.3-0.7)	<.001	0.07 (0.05-0.10)
Other: present	582 (14.3)	411 (26.7)	2.2 (1.9-2.5)	<.001	0.13 (0.10-0.16)
Pattern unable to determine: present	203 (5.0)	160 (10.4)	2.2 (1.8-2.7)	<.001	0.10 (0.08-0.12)
Organized					0.19 (0.16-0.22)
No	1593 (39.2)	1007 (65.3)	3.3 (2.9-3.7)	<.001	
Yes	2304 (56.7)	445 (28.9)	1 [Reference]	NA	
Unknown	165 (4.1)	89 (5.8)	2.8 (2.1-3.7)	<.001	
Contour symmetry					0.37 (0.32-0.42)
Two axes	1876 (46.2)	398 (25.9)	1 [Reference]	NA	
One axis	981 (24.2)	313 (20.4)	1.5 (1.3-1.8)	<.001	
None	1173 (28.9)	788 (51.2)	3.2 (2.7-3.7)	<.001	
Unable to determine	29 (0.7)	39 (2.5)	6.3 (4.0-9.9)	<.001	
Pattern symmetry					0.37 (0.32-0.41)
Two axes	1450 (35.7)	189 (12.3)	1 [Reference]	NA	
One axis	1002 (24.7)	313 (20.4)	2.4 (1.9-3.0)	<.001	
None	1569 (38.7)	1005 (65.3)	4.9 (4.1-5.8)	<.001	
Unable to determine	38 (0.9)	31 (2.0)	6.3 (3.6-10.8)	<.001	
Architectural disorder					0.43 (0.39-0.48)
None	2115 (52.1)	379 (24.6)	1 [Reference]	NA	
Mild	1435 (35.4)	556 (36.2)	2.2 (1.9-2.5)	<.001	
Marked	509 (12.5)	603 (39.2)	6.6 (5.6-7.8)	<.001	
Borders					0.16 (0.13-0.19)
0	2063 (50.8)	486 (31.6)	1 [Reference]	NA	
1	299 (7.4)	114 (7.4)	1.6 (1.3-2.1)	<.001	
2	385 (9.5)	165 (10.7)	1.8 (1.5-2.2)	<.001	
3	300 (7.4)	127 (8.3)	1.8 (1.4-2.3)	<.001	
4	221 (5.4)	148 (9.6)	2.8 (2.3-3.6)	<.001	
5	120 (3.0)	88 (5.7)	3.1 (2.3-4.2)	<.001	NA
6	130 (3.2)	100 (6.5)	3.3 (2.5-4.3)	<.001	
7	87 (2.1)	52 (3.4)	2.5 (1.8-3.6)	<.001	
8	343 (8.5)	151 (9.8)	1.9 (1.5-2.3)	<.001	
Unable to determine	111 (2.7)	107 (6.9)	4.1 (3.1-5.4)	<.001	
Colors					
Light brown	3677 (90.5)	1307 (84.8)	0.6 (0.5-0.7)	<.001	0.28 (0.24-0.32)
Dark brown	3333 (82.0)	1212 (78.7)	0.8 (0.7-0.9)	.004	0.40 (0.35-0.44)
White	698 (17.2)	468 (30.4)	2.1 (1.8-2.4)	<.001	0.20 (0.16-0.23)
Gray	710 (17.5)	304 (19.7)	1.2 (1.0-1.3)	.05	0.10 (0.08-0.13)
Blue	421 (10.4)	291 (18.9)	2.0 (1.7-2.4)	<.001	0.21 (0.17-0.24)
Black	938 (23.1)	572 (37.1)	2.0 (1.7-2.2)	<.001	0.36 (0.31-0.41)
Red	835 (20.6)	514 (33.4)	1.9 (1.7-2.2)	<.001	0.36 (0.31-0.41)
Blue or gray	675 (16.6)	398 (25.8)	1.7 (1.5-2.0)	<.001	0.15 (0.12-0.18)
Blue or white	327 (8.1)	238 (15.4)	2.1 (1.7-2.5)	<.001	0.17 (0.14-0.21)

(continued)

Table 3. Association Between Dermoscopic Criteria With Melanoma Status (continued)

Dermoscopic Criterion	No. (%) of Lesions		OR (95% CI)	*P* Value	ICC (95% CI)[a]
	Nevus (n = 4064)	Melanoma (n = 1541)			
Total colors			1.4 (1.3-1.5)	<.001	0.36 (0.31-0.40)
1	340 (8.4)	78 (5.1)			
2	1373 (33.8)	344 (22.3)			
3	1344 (33.1)	463 (30.1)			
4	678 (16.7)	348 (22.6)			
5	229 (5.6)	171 (11.1)	NA	NA	NA
6	68 (1.7)	84 (5.5)			
7	21 (0.5)	34 (2.2)			
8	6 (0.2)	11 (0.7)			
9	4 (0.1)	8 (0.5)			
Network					
None	1155 (28.4)	496 (32.2)	2.5 (2.1-3.0)	<.001	0.39 (0.34-0.43)
Typical	1057 (26.0)	181 (11.8)	1 [Reference]	NA	0.19 (0.16-0.23)
Atypical	1560 (38.4)	756 (49.1)	2.8 (2.4-3.4)	<.001	0.21 (0.17-0.25)
Both	292 (7.2)	108 (7.0)	2.2 (1.6-2.8)	<.001	0.11 (0.08-0.13)
Network					
Pseudo: present	161 (4.0)	57 (3.7)	0.9 (0.7-1.3)	.65	0.07 (0.05-0.09)
Negative: present	204 (5.0)	107 (6.9)	1.4 (1.1-1.8)	.005	0.15 (0.12-0.18)
Target: present	122 (3.0)	30 (2.0)	0.6 (0.4-1.0)	.03	0.06 (0.05-0.08)
Structureless areas: present	1934 (47.6)	877 (56.9)	1.5 (1.3-1.6)	<.001	0.08 (0.06-0.10)
Hypopigmented areas: present	1244 (30.6)	618 (40.1)	1.5 (1.3-1.7)	<.001	0.17 (0.14-0.20)
Blotch					
Regular: present	374 (9.2)	67 (4.4)	0.4 (0.3-0.6)	<.001	0.08 (0.06-0.10)
Irregular: present	1037 (25.5)	615 (39.9)	1.9 (1.7-2.2)	<.001	0.18 (0.14-0.21)
Blue-white veil: present	759 (18.7)	537 (34.9)	2.3 (2.0-2.7)	<.001	0.34 (0.30-0.39)
Blue-gray granules: present	348 (8.6)	164 (10.6)	1.3 (1-1.5)	.02	0.11 (0.08-0.14)
Scar: present	277 (6.8)	233 (15.1)	2.4 (2.0-2.9)	<.001	0.20 (0.16-0.24)
Peripheral brown dots: present	366 (9.0)	195 (12.7)	1.5 (1.2-1.8)	<.001	0.04 (0.03-0.06)
Blue-gray dots: present	341 (8.4)	172 (11.2)	1.4 (1.1-1.7)	.001	0.16 (0.13-0.19)
Streaks: present	761 (18.7)	402 (26.1)	1.5 (1.3-1.8)	<.001	0.21 (0.17-0.24)
Pseudopods: present	296 (7.3)	215 (14.0)	2.1 (1.7-2.5)	<.001	0.23 (0.19-0.27)
Structures					
White shiny: present	84 (2.1)	78 (5.1)	2.5 (1.8-3.5)	<.001	0.16 (0.13-0.19)
Rhomboid: present	74 (1.8)	16 (1.0)	0.6 (0.3-1.0)	.04	0.05 (0.03-0.06)
Regression: present	391 (9.6)	275 (17.9)	2.0 (1.7-2.4)	<.001	0.11 (0.08-0.13)
Dots					
Regular black: present	123 (3.0)	40 (2.6)	0.9 (0.6-1.2)	.39	0.05 (0.03-0.07)
Regular brown: present	494 (12.2)	98 (6.4)	0.5 (0.4-0.6)	<.001	0.06 (0.04-0.08)
Irregular black: present	392 (9.7)	245 (15.9)	1.8 (1.5-2.1)	<.001	0.13 (0.10-0.16)
Irregular brown: present	854 (21.0)	413 (26.8)	1.4 (1.2-1.6)	<.001	0.12 (0.09-0.14)
Irregular blue: present	116 (2.9)	65 (4.2)	1.5 (1.1-2.0)	.01	0.06 (0.04-0.08)
Irregular red: present	59 (1.5)	34 (2.2)	1.5 (1.0-2.3)	.05	0.06 (0.04-0.08)
Globules					
Regular black: present	76 (1.9)	33 (2.1)	1.1 (0.8-1.7)	.51	0.05 (0.03-0.07)
Regular brown: present	558 (13.7)	121 (7.9)	0.5 (0.4-0.7)	<.001	0.17 (0.13-0.20)
Regular blue: present	45 (1.1)	10 (0.7)	0.6 (0.3-1.2)	.12	0 (0-0.01)
Irregular black: present	286 (7.0)	191 (12.4)	1.9 (1.5-2.3)	<.001	0.14 (0.11-0.17)
Irregular brown: present	786 (19.3)	326 (21.2)	1.1 (1.0-1.3)	.13	0.11 (0.08-0.13)
Irregular blue: present	143 (3.5)	113 (7.3)	2.2 (1.7-2.8)	<.001	0.07 (0.05-0.09)
Vessels					
None	3260 (80.2)	1000 (64.9)	0.5 (0.4-0.5)	<.001	0.46 (0.42-0.51)
Comma	236 (5.8)	40 (2.6)	0.4 (0.3-0.6)	<.001	0.44 (0.40-0.49)
Atypical	293 (7.2)	293 (19.0)	3.0 (2.5-3.6)	<.001	0.26 (0.22-0.30)
Pink veil	251 (6.2)	221 (14.3)	2.5 (2.1-3.1)	<.001	0.15 (0.12-0.18)
Polymorphous	115 (2.8)	127 (8.2)	3.1 (2.4-4.0)	<.001	0.16 (0.13-0.19)

Abbreviations: ICC, intraclass correlation coefficient; NA, not applicable; OR, odds ratio.

[a] The ICC (95% CI) values were added as a measure of interobserver agreement.

Table 4. Measures of Diagnostic Accuracy for 6 Dermascopic Algorithms

Measure	7-Point Checklist (Cut Point ≥3)	CASH (Cut Point ≥6)	Menzies Method	ABCD Rule (TDS Score >4.75)	3-Point Checklist	Chaos and Clues
Sensitivity, % (95% CI)	70.6 (61.5-78.6)	77.9 (69.7-85.1)	95.1 (89.0-98.4)[a]	74.8 (66.0-82.3)	68.9 (59.8-77.1)	82.4 (66.1-96.5)
Specificity, % (95% CI)	57.5 (52.2-62.7)	50.9 (45.4-56.4)	24.8 (20.1-30.1)[b]	59.4 (54.0-64.6)	58.7 (53.4-63.8)	40.2 (35.1-45.5)[c]
ROC area (95% CI)	0.65 (0.59-0.69)	0.65 (0.59-0.69)	0.60 (0.57-0.63)	0.66 (0.62-0.72)	0.64 (0.59-0.69)	0.66 (0.63-0.70)

Abbreviations: CASH, color, architecture, symmetry, and homogeneity; ROC, receiver operating characteristic; TDS, total dermatoscopy score.

[a] Sensitivity of the Menzies method was significantly higher than any other algorithm.

[b] Specificity of the Menzies method was significantly lower than any other algorithm.

[c] Specificity of chaos and clues was significantly lower than the 7-point checklist, the 3-point checklist, and the ABCD rule.

Comparison of Diagnostic Accuracy of the 6 Simplified Algorithms

Measures of diagnostic accuracy for the ABCD rule, the Menzies method, the 7-point checklist, the 3-point checklist, chaos and clues, and CASH are given in **Table 4.** Note that this analysis was artificially constructed by using the participants' consensus evaluation of individual criteria (ie, when ≥50% of the participants identified a dermoscopic feature for a given study lesion, the attribute was considered present) and that participants did not directly score algorithms in a head-to-head comparison scenario. For these analyses, the data are presented with defined cut points for melanoma diagnosis. The Menzies method had the highest sensitivity for melanoma detection (95.1%; 95% CI, 89.0%-98.4%), significantly higher than any other method ($P < .001$), and the 3-point checklist had the lowest (68.9%; 95% CI, 59.8%-77.1%). The ABCD rule had the highest specificity (59.4%; 95% CI, 54.0%-64.6%), which was significantly higher compared with chaos and clues (40.2%; 95% CI, 35.1%-45.5%) and the Menzies method, which had the lowest (24.8%; 95% CI, 20.1%-30.1%) compared with any other ($P < .001$). Chaos and clues had significantly lower specificity compared with the ABCD rule and the 3- and 7-point checklists. The **Figure** shows the ROC curves of the 6 algorithms. No significant differences in ROC areas were observed in CASH, the 7-point checklist, the 3-point checklist, chaos and clues, and the ABCD rule ($P = .44$). However, the Menzies method had a lower ROC area compared with CASH, the 7-point checklist, the 3-point checklist, the ABCD rule, and chaos and clues, with P values for each comparison of .03, .03, .007, .001, and <.001, respectively.

Discussion

In this study, which involved participants of varied backgrounds who reported comfort with and regular use of dermoscopy, we revalidated the diagnostic importance of well-described criteria associated with melanoma, such as atypical network, irregular blotch, regression, streaks, pseudopods, atypical dots or globules, atypical vessels, any blue or white color, and blue-white veil. However, we found that these criteria had poor to fair levels of interobserver agreement. Criteria with the highest levels of discriminatory power and interobserver agreement included features not always highlighted in existing algorithms, such as comma vessels and absence of vessels, as well as subjective features that quantify the overall organization of a lesion, namely, architectural disorder and symmetry of pattern and contour. We further found that 6 simplified dermoscopy algorithms had similar but modest levels of diagnostic accuracy.

Few reproducibility studies of dermoscopic features have been performed, particularly investigating the discriminatory power and interobserver and intraobserver agreement of specific criteria. An Internet consensus meeting of dermoscopy experts in 2003 found that pattern analysis, the ABCD rule, the 7-point checklist, and the Menzies method all have high sensitivity and specificity for the diagnosis of melanoma.[11] However, the interobserver agreement of the diagnostic methods was moderate, and many individual diagnostic structures had poor levels of interobserver agreement. The authors suggested that this discrepancy might be attributable to the importance of the overall dermoscopic gestalt of a given lesion to the assignment of a final diagnosis, independent of the recognition of individual criteria.[11] Indeed, experts usually do not apply algorithms. In other words, evaluators may assign a diagnosis based on the overall impression of a lesion and then search for criteria to fit their decision. To avoid this potential bias, participants in our study evaluated the presence and absence of dermoscopic features but did not apply an algorithm or make a diagnosis. A comparative study[8] of pattern analysis and the different algorithms among nonexperts have also found generally poor interobserver agreement for most individual dermoscopic criteria but much better results for the method as a whole. This interpretation is supported by a study[12] of dermatology residents that found that pattern analysis, defined by the authors as the "simultaneous assessment of the diagnostic value of all dermoscopy features shown by the lesion,"[12(p 981)] had a higher diagnostic accuracy compared with the ABCD rule of dermoscopy and the 7-point checklist.

Of interest, in the present study, several features that indicate overall organization and symmetry had the highest agreement and discriminatory power, such as architectural disorder, contour asymmetry, and dermoscopic pattern asymmetry. These concepts have previously been summarized as disarrangement in appearance or chaos and support the usefulness of chaos and clues[7] and the 3-point checklist,[13] which were created for use in melanocytic and nonmelanocytic lesions. Reassuringly, well-designed, prospective clinical studies[7,8,14,15] have found that use of dermoscopy significantly improves the ability of general practitioners to evaluate pigmented lesions in the primary care setting. Indeed, the 3-point checklist was tested in a clinical setting and allowed

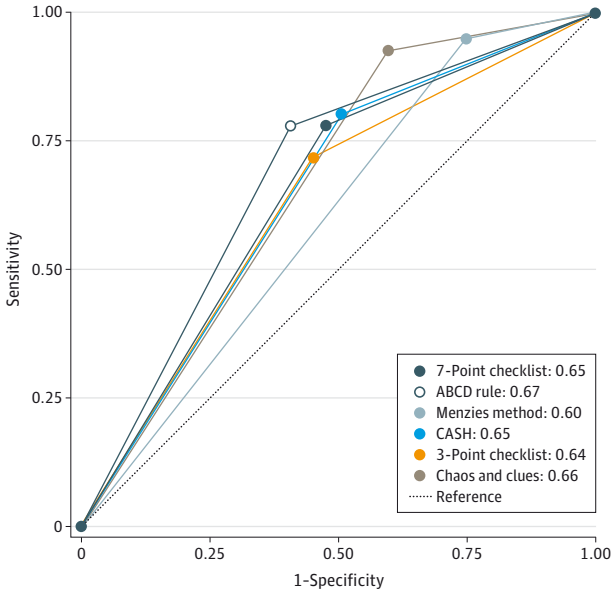
primary care physicians to perform 25.1% better triage of skin lesions suggestive of skin cancer compared with naked-eye examination alone.[14] However, it remains unknown how general practitioners or novices rely on overall dermoscopic gestalt vs application of a dermoscopic algorithm when using dermoscopy in the daily clinical setting. To more broadly promote the use of dermoscopy in the primary care setting, our results suggest that significant efforts are needed to standardize and improve dermoscopic terminology, which is one of the central goals of the International Skin Imaging Collaboration Melanoma Project.[16,17]

Our data suggest that features that quantify the overall organization of a lesion (eg, architectural disorder and pattern asymmetry) have higher levels of interobserver agreement and discriminatory power than many well-known dermoscopic structures (eg, atypical network or irregular blotch); thus, criteria for overall organization of a lesion may not be sufficiently emphasized in dermoscopic algorithms for melanoma diagnosis. Specific dermoscopic structures with low prevalence, such as negative network, may still be robust criteria for melanoma diagnosis but had poor agreement and low discriminatory power in this study because participants may have received insufficient training to accurately identify them. Accordingly, criteria that are useful in melanoma diagnosis should not be abandoned but rather readdressed and potentially refined through further study. This point also highlights the evolving nature and current lack of standardization of dermoscopy teaching worldwide and the critical need to determine effective teaching methods of dermoscopy.

Several factors may contribute to the poor interobserver agreement levels observed in this study. First, participants may not have received sufficient training in the definitions of criteria or, despite training, they used different definitions of criteria, potentially influenced by their personal experience with dermoscopy. To help mitigate these potential factors, we created algorithm tutorials with definitions of criteria. However, completion of tutorials was not required for participation. Second, the interobserver agreement levels may reflect the range of expertise levels of participants in that certain criteria require significant training for mastery. Third, a participant's gestalt diagnosis of a lesion may have affected their criteria selection; if so, a participant may have preferentially assigned some criteria and ignored others. Lastly, criteria may simply be inherently unreliable. For this point, it is important to recognize that tests in medicine are frequently subject to limitations in human judgment and generally do not exceed fair levels of interobserver agreement. In addition, interpretation of the ICC as levels of agreement among reviewers has limitations. When the ICC is high, we can be assured that the agreement level for a given attribute is good. However, a low ICC may be attributable to a suboptimally designed evaluation process. For example, small technical differences in imaging, such as variations in focus or contrast, can have large effects on measure of agreement. In addition, evaluations were performed online, and users viewed images under noncalibrated conditions (eg, variable image display monitors and room lighting).

There are multiple limitations of this study. First, there was a relatively low rate of study completion with likely participation bias for more experienced dermoscopists. As a result, our results may not be generalizable to beginners. Second, we assessed diagnostic accuracy through the artificial scenario of a reader study, which may not be representative of decisions made during live patient examinations. Third, the image data set was not representative of the entire spectrum of melanocytic lesions because it excluded facial, acral, and amelanotic lesions and was biased toward diagnostically challenging lesions with few banal nevi included. In addition, nonmelanocytic lesions were excluded, and the study assumes that participants would apply these criteria after reliably identifying lesions as melanocytic in origin (ie, 2-step algorithm). Thus, comparison of measures of diagnostic accuracy for the included algorithms may not accurately reflect real-life sensitivities and specificities. Finally, diagnostic performance of algorithms was assessed based on consensus evaluations (≥50%) for individual criteria and not directly by individual participants or experts.

Figure. Comparison of the Diagnostic Accuracy of the Dermoscopic Algorithms



Receiver operating characteristic curves for 6 dermoscopic algorithms were evaluated. CASH indicates color, architecture, symmetry, and homogeneity.

Conclusions

Algorithms are generally well accepted to be helpful in training novices in discriminating processes. Therefore, the criteria of an ideal algorithm should be easy to learn, valid, and reliable. Unfortunately, to our knowledge, no dermoscopic algorithm has emerged with these characteristics for melanoma recognition. Our results confirm the need to further improve dermoscopic terminology, criteria, and algorithms. To do so, future studies may benefit from crowd-sourcing and collective intelligence approaches,[18] as well as the public image archive being created in the International Skin Imaging

Collaboration Melanoma Project, which permits analysis and comparison of the areas within a lesion that users select as having unique dermoscopic structures.[16,17] We hope these efforts will lead to a unified dermoscopy algorithm, automated detection of criteria, and clinical decision support systems that facilitate population-based melanoma screening efforts.[19]

ARTICLE INFORMATION

Accepted for Publication: February 18, 2016.

Published Online: April 13, 2016. doi:10.1001/jamadermatol.2016.0624.

Author Affiliations: Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York (Carrera, Marchetti, Dusza, Halpern, Marghoob); Melanoma Unit, Department of Dermatology, Hospital Clinic Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Centro de Investigacion Biomedica en red de enfermedades raras, Barcelona, Spain (Carrera, Malvehy, Puig); Dermatology Unit, Second University of Naples, Naples, Italy (Argenziano); Department of Dermatology, University Hospital Zürich, Zürich, Switzerland (Braun); Dermatology Service, Aurora Skin Cancer Center, Universidad Pontificia Bolivariana, Medellín, Colombia (Jaimes); Department of Dermatology, Medical University of Vienna, Vienna, Austria (Kittler); Sydney Melanoma Diagnostic Centre, Sydney Cancer Centre, Royal Prince Alfred Hospital, Camperdown, Australia (Menzies); Discipline of Dermatology, The University of Sydney, New South Wales, Australia (Menzies); Center for Environmental, Genetic, and Nutritional Epidemiology, Department of Diagnostic, Clinical, and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy (Pellacani); Skin and Cancer Associates, Plantation, Florida (Rabinovitz); Department of Dermatology, Sheba Medical Center, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel (Scope); Dermatology Research Centre, The University of Queensland, Brisbane, Queensland, Australia (Soyer); School of Medicine, Translational Research Institute, Brisbane, Queensland, Australia (Soyer); Clinic for Dermatology, Allergology, and Environmental Medicine, Klinik Thalkirchner Straße Städt, Klinikum München GmbH, Munich, Germany (Stolz); Department of Dermatology, Medical University of Graz, Graz, Austria (Hofmann-Wellenhof, Zalaudek).

Author Contributions: Drs Carrera and Marghoob had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Carrera and Marchetti are co-first authors.
Study concept and design: Dusza, Braun, Halpern, Jaimes, Malvehy, Puig, Scope, Hofmann-Wellenhof, Marghoob.
Acquisition, analysis, or interpretation of data: Carrera, Marchetti, Dusza, Argenziano, Braun, Jaimes, Kittler, Malvehy, Menzies, Pellacani, Rabinovitz, Scope, Soyer, Stolz, Hofmann-Wellenhof, Zalaudek, Marghoob.
Drafting of the manuscript: Carrera, Marchetti, Dusza.
Critical revision of the manuscript for important intellectual content: Carrera, Marchetti, Dusza, Argenziano, Braun, Halpern, Jaimes, Kittler, Malvehy, Menzies, Pellacani, Puig, Rabinovitz, Scope, Soyer, Stolz, Hofmann-Wellenhof, Zalaudek, Marghoob.
Statistical analysis: Dusza.
Administrative, technical, or material support: Kittler, Stolz, Zalaudek, Marghoob.
Study supervision: Dusza, Halpern, Marghoob.

Conflict of Interest Disclosures: Dr Halpern reported receiving consulting fees from Canfield Scientific Inc, DermTech, and SciBase. Dr Rabinovitz reported serving as a clinical investigator for 3 Gen LLC and Canfield and as a speaker for 3 Gen LLC. Dr Soyer reported receiving support in part from Australian National Health and Medical Research Council Practitioner Fellowship APP1020145 and being a shareholder of e-derm consult GmbH and MoleMap by Dermatologists Ltd Pty. He provides teledermatologic reports regularly for both companies. Dr Hofmann-Wellenhof reported being a shareholder of e-derm consult GmbH. He provides teledermatologic reports regularly for this company. No other disclosures were reported.

Funding/Support: This research was funded in part through National Institute of Health/National Cancer Institute Cancer Center Support Grant P30 CA008748. The research at the Melanoma Unit in Barcelona is partially funded by grants 12/00840 and PI15/00716 from Fondo de Investigaciones Sanitarias, the CIBER de Enfermedades Raras of the Instituto de Salud Carlos III, the AGAUR 2014_SGR_603 of the Catalan Government, a grant from Fundació La Marató de TV3, 201331-30, and grant CE_CIP-ICT-PSR-13-7 from the European Commission under the 7th Framework Programme (Diagnoptics).

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Additional Contributions: We are in debt to the participants and members of the International Dermoscopy Society who completed the study; Al Kopf, MD, who contributed his knowledge and personal collection of images; and Gerald Gabler, MSc, who developed and created the study website.

REFERENCES

1. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol*. 2002;3(3):159-165.

2. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol*. 1998;134(12):1563-1570.

3. Stolz W, Riemann A, Cognetta AB, et al. ABCD rule of dermoscopy: a new practical method for early recognition of malignant melanoma. *Eur J Dermatol*. 1994;4(7):521-527.

4. Soyer HP, Argenziano G, Zalaudek I, et al. Three-point checklist of dermoscopy: a new screening method for early detection of melanoma. *Dermatology*. 2004;208(1):27-31.

5. Henning JS, Dusza SW, Wang SQ, et al. The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy. *J Am Acad Dermatol*. 2007;56(1):45-52.

6. Menzies SW, Ingvar C, Crotty KA, McCarthy WH. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. *Arch Dermatol*. 1996;132(10):1178-1182.

7. Rosendahl C, Cameron A, McColl I, Wilkinson D. Dermatoscopy in routine practice: 'chaos and clues'. *Aust Fam Physician*. 2012;41(7):482-487.

8. Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. *Arch Dermatol*. 2005;141(8):1008-1014.

9. Pizzichetta MA, Talamini R, Marghoob AA, et al. Negative pigment network: an additional dermoscopic feature for the diagnosis of melanoma. *J Am Acad Dermatol*. 2013;68(4):552-559.

10. Balagula Y, Braun RP, Rabinovitz HS, et al. The significance of crystalline/chrysalis structures in the diagnosis of melanocytic and nonmelanocytic lesions. *J Am Acad Dermatol*. 2012;67(2):194.e1-194.e8.

11. Argenziano G, Soyer HP, Chimenti S, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *J Am Acad Dermatol*. 2003;48(5):679-693.

12. Carli P, Quercioli E, Sestini S, et al. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. *Br J Dermatol*. 2003;148(5):981-984.

13. Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. *J Am Acad Dermatol*. 2011;64(6):1068-1073.

14. Argenziano G, Puig S, Zalaudek I, et al. Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer. *J Clin Oncol*. 2006;24(12):1877-1882.

15. Menzies SW, Emery J, Staples M, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. *Br J Dermatol*. 2009;161(6):1270-1277.

16. International Society for Digital Imaging of the Skin. http://isdis.net/isic-project/. Accessed October 7, 2015.

17. International Skin Imaging Collaboration (ISIC). Melanoma Project. *ISIC Archive*. https://isic-archive.com/. Accessed October 7, 2015.

18. Kurvers RH, Krause J, Argenziano G, Zalaudek I, Wolf M. Detection accuracy of collective intelligence assessments for skin cancer diagnosis. *JAMA Dermatol*. 2015;151(12):1346-1353.

19. Katalinic A, Waldmann A, Weinstock MA, et al. Does skin cancer screening save lives? an observational study comparing trends in melanoma mortality in regions with and without screening. *Cancer*. 2012;118(21):5395-5402.